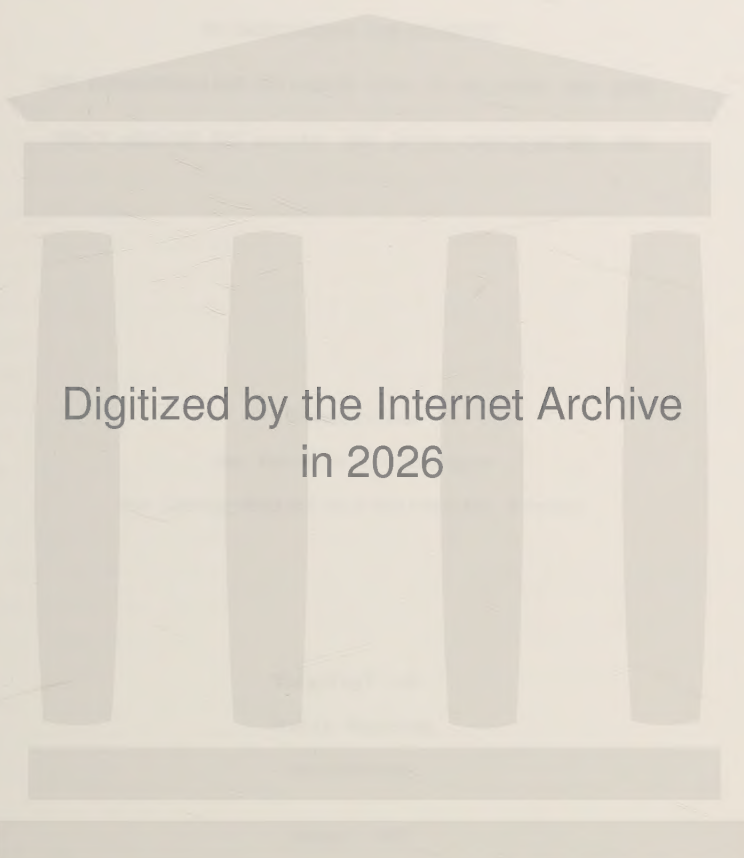


UNTERSUCHUNGEN ZUR STRUKTUR
DES BASALMEMBRANEN-KOLLAGENS
(TYP IV KOLLAGEN) DER MAUS
DURCH ANALYSE VON $\alpha 1(\text{IV})$ -
UND $\alpha 2(\text{IV})$ -SPEZIFISCHER cDNA

Ursula Magdolen

Disbui. München

1987



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41554617_0031

UNTERSUCHUNGEN ZUR STRUKTUR
DES BASALMEMBRANEN-KOLLAGENS (TYP IV KOLLAGEN) DER MAUS
DURCH ANALYSE VON $\alpha 1$ (IV)- UND $\alpha 2$ (IV)-SPEZIFISCHER cDNA

Dissertation
der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

Vorgelegt von
Ursula Magdolen
aus München

Januar 1987

1. Berichterstatter: Prof. Dr. W. Bandlow

2. Berichterstatter: Prof. Dr. K. Wolf

Sondervotum: Prof. Dr. K. Kuhn

Tag der mündlichen Prüfung: 31. 3. 1987

DISSERTATIONS- UND FOTODRUCK FRANK GmbH
8000 München 2, Gabelsbergerstr. 15, Tel. 288663

Aus aktuellen Gründen wurden Teile dieser Arbeit bereits veröffentlicht bzw. sind in Vorbereitung:

Oberbäumer, I., Laurent, M., Schwarz, U., Sakurai, Y., Yamada, Y., Vogeli, G., Voss, T., Siebold, B., Glanville, R. W. und Kühn K. (1985)

"Amino acid sequence of the non-collagenous globular domain (NC1) of the $\alpha 1(IV)$ chain of basement membrane collagen as derived from complementary DNA"
Eur. J. Biochem. 147, 217-224

Schwarz, U., Schuppan, D., Oberbäumer, I., Glanville, R. W., Deutzmann, R., Timpl, R. und Kühn K. (1986)

"Structure of mouse type IV collagen. Amino acid sequence of the C-terminal 511-residue-long triple helical segment of the $\alpha 2(IV)$ chain and its comparison with the $\alpha 1(IV)$ chain"
Eur. J. Biochem. 157, 49-56

Schwarz-Magdolen, U., Oberbäumer I. und Kühn K. (1986)

"cDNA and protein sequence of the NC1 domain of the $\alpha 2$ -chain of collagen IV and its comparison with $\alpha 1(IV)$ "
FEBS Lett. 208, 203-207

Oberbäumer, I. and Schwarz-Magdolen, U.

"Structure of the 3' noncoding regions of $\alpha 1(IV)$ and $\alpha 2(IV)$ "
in Vorbereitung

Inhaltsverzeichnis

<u>A. EINLEITUNG</u>	1
<u>B. AUFGABENSTELLUNG</u>	12
<u>C. ERGEBNISSE</u>	13
Teil I: <u>Isolierung und Charakterisierung von Typ IV-spezifischen Klonen</u>	13
1. Einführung	13
2. Typ IV-spezifische Oligonukleotide und Analyse ihres Hybridisierungs- verhaltens in Northernhybridisierungen	18
3. Isolierung $\alpha 1(IV)$ -spezifischer cDNA-Klone	29
3.1. Identifizierung und Charakterisierung der Klone p5 und p9	29
3.2. Charakterisierung von Klon pFAC	36
4. Identifizierung $\alpha 2(IV)$ -spezifischer cDNA-Klone	38
4.1. Isolierung und Charakterisierung der cDNA-Klone pAIIa und pAIIb	38
4.2. Isolierung und Charakterisierung von Klon pAIIc	45
Teil II: <u>Charakterisierung der NC1-Domäne und der Tripelhelix</u> <u>auf Protein- und DNA-Ebene</u>	48
1. Die Struktur der NC1-Domäne	48
1.1 DNA-Sequenzen und abgeleitete Proteinsequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$	48
1.2 Homologie von $\alpha 1(IV)$ und $\alpha 2(IV)$ im Bereich der NC1-Domäne	53
1.3 Interne Homologie der NC1-Domäne und Homologie zu den Carboxypropeptiden der interstitiellen Kollagene	55

2. Die Struktur der Tripelhelix	59
2.1. DNA-Sequenzen und abgeleitete Proteinsequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$	59
2.2. Vergleich der Proteinsequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$	64
3. Analyse der codierenden cDNA-Sequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$	72
3.1. Sequenzhomologien der cDNA von $\alpha 1(IV)$ und $\alpha 2(IV)$	72
3.2. Nukleotidzusammensetzung und Codonauswahl	75
Teil III: <u>Aufklärung der 3'-nichttranslatierten Sequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$ und Charakterisierung der mRNAs</u>	82
1. Kurze Einleitung	82
2. Vervollständigung der 3'-nichttranslatierten Sequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$	87
2.1. Isolierung und Charakterisierung der $\alpha 1(IV)$ -spezifischen Klone pAI4, pAI5 und pAI6	87
2.2. Identifizierung des 3'-Endes der nichttranslatierten Region von $\alpha 2(IV)$	93
3. Die Sequenzen der 3'-nichttranslatierten Regionen von $\alpha 1(IV)$ und $\alpha 2(IV)$	95
4. Charakterisierung der Typ IV mRNA	103
D. DISKUSSION	110
E. ZUSAMMENFASSUNG	132
F. MATERIAL UND METHODEN	134
G. LITERATUR	155
Abkürzungen	

A. EINLEITUNG

Basalmembranen stellen eine spezialisierte Form extrazellulärer Matrix dar, die als dünne amorphe Struktur im Organismus weit verbreitet ist. In elektronenmikroskopischen Aufnahmen erscheinen die 30 bis 300 nm dicken Basalmembranen gewöhnlich zwei- oder dreischichtig. Eine elektronendichtere Mittelschicht, die Lamina densa, wird von ein oder zwei elektronenoptisch dünneren Schichten, den Lamina lucida, flankiert (Abb.1).

Viele verschiedene Zelltypen wie epitheliale und endotheliale Zellen, Muskelzellen, Fettzellen und Schwannzellen synthetisieren Basalmembranen und schaffen sich dadurch eine Abgrenzung zur umgebenden interstitiellen Matrix oder zu anderen Zelltypen. Die daraus resultierende Kompartimentierung der verschiedenen Gewebe sorgt für die Aufrechterhaltung der Gewebestrukturen vor allem bei der Regeneration nach Verletzungen und bei der Morphogenese (Vracko, 1982). Basalmembranen sind Stützstrukturen und verhindern als mechanische Barrieren den Durchtritt von Zellen und Makromolekülen. Ihre Fähigkeit der selektiven Filtration ist besonders wichtig für den Stoffaustausch in Nierenglomeruli und Kapillaren (Farquar, 1982). Spezialisierte Basalmembranen finden sich in neuromuskulären Synapsen, wo sie die Lokalisation spezifischer Komponenten der Endplatte kontrollieren (Burden et al., 1979; Sanes et al., 1978). Somit sind sie maßgeblich an der Signalübertragung sowie an der Regeneration von Nerv-Muskel-Kontakten beteiligt. Interaktionen von Zellen mit Basalmembranen spielen eine Rolle bei Prozessen wie Polarisation z.B. von Epithelzellen, Zellmigration und -adhäsion, Proliferation und Differenzierung (Timpl und Dziadek, 1986). Diese Wechselwirkungen sind daher insbesondere in der Embryonalphase von großer Bedeutung, aber auch für die Aufrechterhaltung differenzierter Entwicklungsstadien.

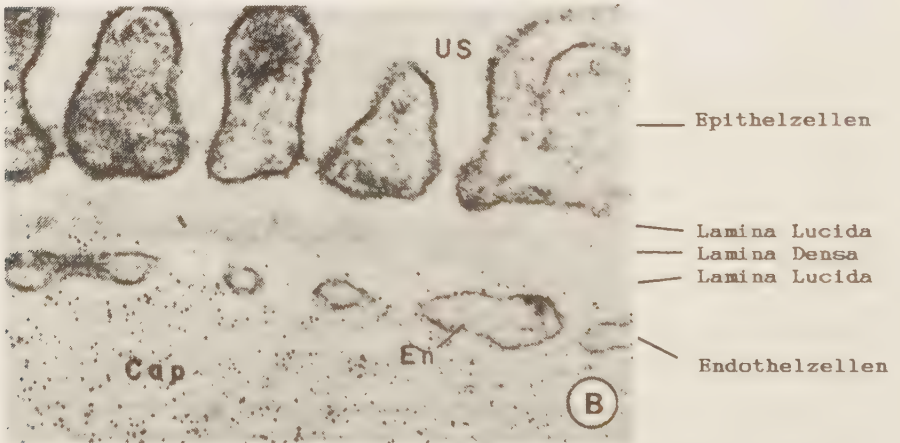


Abb.1

- A. Elektronenmikroskopische Aufnahme einer Basalmembran aus der Kornea eines Hühnerembryos (aus "Molecular Biology of the Cell", 1983; Herausgeber: Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J. D.) Die Basalmembran (BL) trennt epitheliale Zellen (E) von der darunterliegenden interstitiellen Matrix mit ihren Kollagenfibrillen (C).
- B. Elektronenmikroskopische Aufnahme einer Basalmembran aus einem Nierenglomerulus der Ratte (aus "New Trends in Basement Membrane Research", 1982; Herausgeber: Kühn, K., Schoene, H. and Timpl, R.) Die Basalmembran setzt sich aus drei Schichten zusammen: der Lamina densa und den flankierenden Lamina lucida. Eine der beiden Lamina lucida ist der Epithelzellschicht des Harnraumes (US) zugewandt, die andere der Endothelzellschicht des Kapillarraumes (Cap).

Die molekularen Grundlagen der vielfältigen Funktionen der Basalmembranen sind zwar weitgehend unbekannt, jedoch hat es die Charakterisierung der molekularen Hauptbestandteile der Basalmembranen ermöglicht, diesen Komponenten spezifische Funktionen zuzuordnen (Tab.1).

Tabelle 1: Hauptbestandteile der Basalmembran

Protein	Molekulargewicht	Funktionen
Typ IV Kollagen	550-600 KD	Gerüststruktur
Laminin	1000 KD	Zellbindung
Heparansulfatproteoglykan	130 KD	Filtration,
	450 KD	diverse Interaktionen
Entactin	150 KD	
Nidogen	150 KD	Bindung an Laminin
Fibronektin (nur embryonale Basalmembranen?)	450 KD	Zellbindung

Die Hauptmasse der Basalmembranen besteht aus verschiedenen Glykoproteinen. Typ IV Kollagen (Kefalides, 1973), dessen Vorkommen auf Basalmembranen beschränkt ist, bildet das Grundgerüst der Matrix.

Laminin (Timpl et al., 1979), ein Glykoprotein, das aus drei Peptidketten besteht, ist ebenfalls nur in Basalmembranen zu finden. Es hat ein Molekulargewicht von ca. 1 Million Dalton und ist aufgrund seiner Multidomänenstruktur prädestiniert, mehrere biologische Aufgaben zu erfüllen. So besitzt Laminin Bindungsstellen für Heparin und Heparansulfat (Sakashita et al., 1980), und damit für Seitenketten von Heparansulfatproteoglykanen. Laminin vermittelt das Anhaften von endo- und epithelialen Zellen an Typ IV Kollagen (Terranova et al., 1980) und fördert Wachstum und Überleben von Neuronen (Edgar et al., 1984).



Die Heparansulfatproteoglykane, Basalmembrankomponenten mit polyanionischen Seitenketten, wirken vermutlich als Barrieren für geladene Makromoleküle (Farquar, 1982). Die 130 KD Form der Heparansulfatproteoglykane bindet zudem in Affinitätschromatographiestudien an Laminin, Fibronectin und Typ IV Kollagen (Fujiwara et al., 1984).

Als weitere Basalmembrankomponenten konnten Nidogen (Timpl et al., 1983), ein sulfatiertes Protein (Paulsson et al., 1985), das an Laminin bindet (Dziadek und Timpl, 1985), und Entactin (Carlén et al., 1981), das ebenfalls sulfatiert ist, identifiziert werden. Inzwischen ist nachgewiesen, daß Nidogen und Entactin modifizierte Formen ein und desselben Proteins darstellen (Chung, persönliche Mitteilung).

Fibronectin, primär ein Protein des Interstitiums, kommt vermutlich nur in embryonalen Basalmembranen vor (Stenman und Vaheri, 1978; Vartiomaara et al., 1979). Es ist ebenfalls ein Multidomänenprotein mit verschiedenen biologischen Funktionen (Yamada, 1983).

Die Verteilung der einzelnen Proteine zwischen Lamina densa und Lamina lucida ist nicht vollständig geklärt. Es gibt Hinweise darauf, daß Typ IV Kollagen im wesentlichen in der Lamina densa lokalisiert ist (Yaoita et al., 1978), während die anderen Komponenten so eingelagert sind, daß ihre funktionellen Gruppen an der Außenseite der Basalmembran mit anhaftenden Zellen und der interstitiellen Matrix in Wechselwirkung treten können (Courtroy et al., 1982; Kanwar und Farquar, 1979).

Basalmembrankollagen

Eine wichtige Komponente von Basalmembranen ist Typ IV Kollagen, das das Grundgerüst der Matrix bildet. Dieses Protein ist ein Mitglied der umfangreichen Familie der Kollagene. Die Kollagene sind eine Gruppe von Strukturproteinen, die in vielzelligen Organismen einen Anteil von bis zu 25% am Gesamtprotein haben. Bisher konnten 11 verschiedene Kollagene identifiziert werden (Tab.2).

Tabelle 2: Kollagentypen

Typ	Ketten	Vorkommen
I	$\alpha 1, \alpha 2$	Knochen, Sehnen, Haut, Dentin, Faszien, Gefäße, innere Organe
II	$\alpha 1$	Knorpel, Glaskörper
III	$\alpha 1$	Haut, Gefäße, innere Organe
IV	$\alpha 1, \alpha 2$	Basalmembranen
V	$\alpha 1, \alpha 2$	glatte Muskulatur, Plazenta, Kornea, Knochen
VI	$\alpha 1, \alpha 2$	Blutgefäße (Intima), Haut, Gebärmutter, Plazenta
VII		Amnion
VIII	$\alpha 1$	Endothelzellen
IX	$\alpha 1, \alpha 2, \alpha 3$	hyaliner Knorpel
X	$\alpha 1$	epiphysialer Knorpel
XI	$\alpha 1, \alpha 2, \alpha 3$	Knorpel
D (Drosophila)		1. und 2. Instar Larven
N (Nematoden)		Kutikula

Das charakteristische Strukturmerkmal der Kollagene ist die Tripelhelix, die durch Aufwindung dreier Peptidketten, genannt α -Ketten, entsteht. Wichtige Voraussetzung für die Ausbildung dieser Struktur ist, daß jede dritte Aminosäure

der Peptidketten ein Glycin ist. Durch den je nach Kollagentyp variierenden hohen Anteil an Prolin und Hydroxyprolin wird die Tripelhelix zusätzlich stabilisiert. Neben Prolin kann auch die Aminosäure Lysin hydroxyliert und anschließend glykosyliert werden, beides posttranslationale Modifikationen, die typisch für Kollagene sind.

Am besten untersucht sind die interstitiellen Kollagene (Typ I-III), von denen sich Typ IV Kollagen in vieler Hinsicht unterscheidet.

Kennzeichnend für die interstitiellen Kollagene ist ihre Fähigkeit, zu Fibrillen zu aggregieren. Die einzelnen Kollagenmoleküle lagern sich aneinander, und zwar parallel versetzt (Abb.2). Hierfür ist die Unterteilung der Tripelhelices in vier homologe sogenannte D-Domänen (D1-D4) verantwortlich. In diesen homologen D-Domänen sind polar geladene und hydrophobe Aminosäuren innerhalb der Tripelhelices in der Weise angeordnet, daß eine parallel versetzte Aggregation der Moleküle erzwungen wird. Dadurch werden maximale Wechselwirkungen der interaktiven Aminosäuren benachbarter Moleküle erreicht (Hulmes et al., 1973, Hofmann et al., 1980).

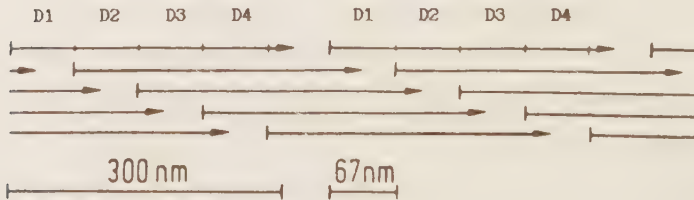


Abb.2: Anordnung der Moleküle in den Fibrillen der interstitiellen Kollagene
Die Kollagenmoleküle lagern sich parallel aneinander, und zwar versetzt um die Einheit D.

Die interstitiellen Kollagene liegen zunächst in einer Vorläuferform, dem Prokollagen vor. Die Prokollagene besitzen an ihren C- und N-terminalen Enden nichthelikale globuläre Domänen, die eine Aggregation der Moleküle verhindern. Erst nach Abspaltung dieser terminalen Domänen können sich die Moleküle zu Fibrillen zusammenlagern.

Die kollagene Komponente der Basalmembranen, Typ IV Kollagen, unterscheidet sich in mehreren Strukturmerkmalen von den interstitiellen Kollagenen.

Typ IV Kollagen wurde aus Basalmembranen der Linsenkapsel (Gay und Miller, 1979; Dixit und Kang, 1979) und Nierenglomeruli (Trüeb et al., 1982; Dixit, 1980) von Rindern gewonnen, sowie aus menschlicher Plazenta (Glanville et al., 1979) und dem EHS-Tumor (Engelbreth-Holm-Swarm) der Maus (Orkin et al., 1977; Timpl et al., 1979). Erste Aminosäureanalysen von Typ IV Kollagen zeigten, daß seine Zusammensetzung von der der Kollagene Typ I-III abweicht. Aminosäuren wie Arginin und Alanin sind unterrepräsentiert, während der Anteil an Hydroxyprolin und -lysin erhöht ist (Kefalides, 1973). Ein hoher Prozentsatz der Hydroxyllysine ist außerdem glykosyliert, was den vergleichsweise hohen Carbohydratgehalt von 10% erklärt.

Aus Basalmembranen konnte Typ IV Kollagen nur durch Säureextraktion oder mit Hilfe von Proteasen in Bruchstücken solubilisiert werden. Da die kontinuierlichen Tripelhelices der interstitiellen Kollagene proteaseresistent sind, war die Proteasesensitivität von Typ IV Kollagen ein erster Hinweis darauf, daß hier der tripelhelikale Bereich mehrfach unterbrochen ist. In elektronenmikroskopischen Studien einzelner Moleküle waren zudem zahlreiche Knicke in der ansonsten starren Tripelhelix sichtbar (Glanville et al., 1982; Hofmann et al., 1984), deren Lokalisation mit den Spaltmustern nach Pepsinabbau in Einklang gebracht werden konnte.

Elektronenmikroskopische Aufnahmen von Typ IV Kollagenmolekülen zeigten außerdem, daß die tripelhelikale Region länger ist als bei den Typen I-III, nämlich 400 nm gegenüber 300 nm, und daß an einem Ende ein globularer Bereich ausgebildet ist (Timpl et al., 1981; Fessler und Fessler, 1982, Oberbäumer et al., 1982).

Die aufgeführten strukturellen Besonderheiten des Typ IV Kollagen, sowie das Fehlen von Fibrillen in der amorph erscheinenden Basalmembran ließen vermuten, daß Typ IV in einer anderen makromolekularen Struktur organisiert ist als die interstitiellen Kollagene. Elektronenmikroskopische Analysen von partialverdautem und säureextrahiertem Typ IV Kollagen erlaubten schließlich eine Interpretation hinsichtlich der möglichen makromolekularen Organisation. Hier sind tetramere spinnenartige Strukturen sichtbar, die durch eine ca. 30 nm lange Überlappung der Enden von vier Molekülen zustande kommen. Zudem treten Dimere auf, die über ihre globulären Enden verknüpft sind. Der Nachweis dieser beiden Strukturen führte zur Postulierung eines Netzwerkmodells (Timpl et al., 1981). Wie in Abb.3 dargestellt, lagern sich demnach die Moleküle nicht parallel aneinander wie bei den interstitiellen Kollagenen, sondern aggregieren über ihre gleichen Enden. Jeweils zwei Monomere werden über ihre globulären Domänen quervernetzt, während sich am anderen Ende vier Moleküle über ihre sogenannte 7S Domänen lateral anlagern (Risteli et al., 1980; Timpl et al., 1981; Kühn et al., 1981; Oberbäumer et al., 1982). Stabilisiert werden die Aggregate durch Disulfidbrücken und andere nicht reduzierbare kovalente Bindungen. Man nimmt an, daß die mehrfachen Unterbrechungen der Tripelhelix dem entstandenen Netzwerk Flexibilität verleihen.

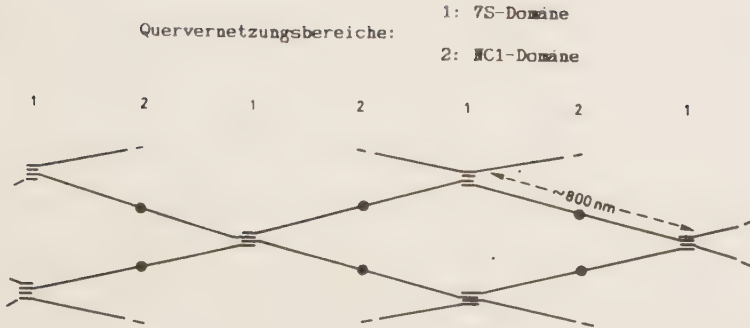


Abb.3: Netzwerkmodell der makromolekularen Struktur von Typ IV Kollagen

Der Aufbau eines Moleküls mit den Quervernetzungsdomänen an beiden Enden ist in Abbildung 4 schematisch dargestellt. Durch Biosynthesestudien (Fessler und Fessler, 1982) und Fragmentanalysen (Dixit et al., 1983) konnte die Orientierung der Moleküle in Bezug auf C- und N-Terminus bestimmt werden.

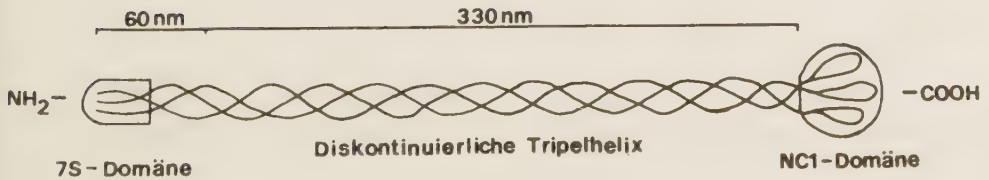


Abb.4: Schematische Darstellung eines Typ IV Kollagenmoleküls mit seinen terminalen Domänen 7S und NC1

Die Trennung der α -Ketten von Typ IV Kollagen aus Zellkulturmedien durch Elektrophorese unter reduzierenden Bedingungen zeigte zwei verschiedene Ketten, $\alpha 1(IV)$ und $\alpha 2(IV)$, mit Molekulargewichten von 185 KD ($\alpha 1$) und 170 KD ($\alpha 2$) (Alitalo et al., 1980; Tryggvason et al., 1980). Vergleich der Peptidmuster beider Ketten und immunologische Analysen belegten, daß es sich um zwei verschiedene Genprodukte handeln muß. Es spricht alles dafür, daß sich die Typ IV Moleküle aus zwei $\alpha 1$ - und einer $\alpha 2$ -Kette zusammensetzen (Dixit und Kang, 1979; Trüb et al., 1982), wenn man auch Homotrimere nicht ausschließen kann. Der Vergleich von Typ IV Kollagen aus Zellkulturmedien mit Typ IV Kollagen aus Matrix ergab keinen Hinweis auf proteolytische Prozessierung, wie sie bei den interstitiellen Kollagenen stattfindet. Somit konnte man davon ausgehen, daß bei der Quervernetzung von Typ IV Kollagen keine, oder nur sehr kurze Propeptide abgespalten werden (Alitalo et al., 1980; Oberbäumer et al., 1982).

Einen hohen Stellenwert bei der Charakterisierung von Typ IV Kollagen hat die Aufklärung der Primärstruktur durch Sequenzanalyse. Abbildung 5 (S. 18) zeigt schematisch die Bereiche aus $\alpha 1(IV)$ und $\alpha 2(IV)$, für die zu Beginn der vorliegenden Arbeit Proteindaten von Mensch und Maus vorlagen. So waren bereits über 50 % der $\alpha 1(IV)$ -Proteinsequenz bekannt. Ein 914 Aminosäuren langes Pepsinfragment aus dem C-terminalen Bereich der Tripelhelix (Mensch) war vollständig sequenziert worden (Babel und Glanville, 1984). Zum Teil existierten auch Daten von Maus und Rind für diese Region (Schuppan et al., 1982; Schuppan et al., 1984). Wenig Informationen dagegen waren über die Primärstruktur der $\alpha 2$ -Kette vorhanden. Die Sequenz des 55 KD Pepsinfragments P2 vom C-Terminus der Tripelhelix (Maus) war zwar zu 90 % aufgeklärt, jedoch konnten mangels Überlappung nicht alle Teilfragmente in der richtigen Reihenfolge angeordnet werden.

Die Sequenzanalyse der NC1-Domane erwies sich als weitaus schwieriger als die der tripelhelikalen Region. Aufgrund des hydrophoben Charakters dieses Bereichs konnten lediglich nach Bromcyanspaltung einzelne Peptide sequenziert werden, deren Anordnung jedoch nicht möglich war.

B. AUFGABENSTELLUNG

Typ IV Kollagen bildet im Gegensatz zu den interstitiellen Kollagenen ein Netzwerk aus, das als Gerüststruktur den Basalmembranen mechanische Stabilität und Flexibilität verleiht. Die Netzwerkpolymerisation erfolgt über die Enden der Typ IV Moleküle, so daß den terminalen Domänen 7S und NC1 eine wichtige funktionelle Bedeutung zukommt.

Nachdem es sich als äußerst schwierig erwies, die NC1-Domäne proteinchemisch zu charakterisieren, sollte die Primärstruktur der C-terminalen Region mit molekularbiologischen Methoden aufgeklärt werden. Weiterhin war von großem Interesse, tripelhelikale Sequenzen der $\alpha 2(IV)$ -Kette zu erhalten, um in diesem Bereich die molekulare Struktur von $\alpha 1(IV)$ und $\alpha 2(IV)$ direkt vergleichen zu können. Die Aufgabenstellung dieser Arbeit war daher, $\alpha 1$ - und $\alpha 2$ -spezifische cDNA-Klone zu isolieren, um im einzelnen folgende Ziele zu verfolgen:

1. Aufklärung der Primärstruktur der globulären Domäne;
2. Vervollständigung der tripelhelikalen Sequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$
3. Homologievergleich von $\alpha 1(IV)$ und $\alpha 2(IV)$;

Homologievergleiche mit interstitiellen Kollagenen

4. Charakterisierung der mRNA von $\alpha 1(IV)$ und $\alpha 2(IV)$

C. ERGEBNISSE

Teil I: Isolierung und Charakterisierung von Typ IV-spezifischen cDNA-Klonen

1. Einführung

Das vorrangige Ziel dieser Arbeit war, die Primärstruktur der NC1-Domäne am C-Terminus von Typ IV Kollagen durch die Isolierung und Analyse der entsprechenden cDNA vollständig aufzuklären.

Proteinsequenzen dieses Bereichs waren für $\alpha 1(IV)$ und $\alpha 2(IV)$ bereits teilweise bekannt (siehe Abb.5, S. 18). Der Ausgangspunkt für die geplante Isolierung von cDNA waren von diesen Partialpeptidsequenzen abgeleitete DNA-Sequenzen. Nach diesen DNA-Sequenzen konnten synthetische Oligonukleotide hergestellt und zur Gewinnung von Typ IV-spezifischer cDNA eingesetzt werden.

Noch in der Anfangsphase dieser Arbeiten ergab es sich glücklicherweise, daß von M. Kurkinen (Rutgers, U.S.A.) erste cDNA-Sequenzdaten der NC1-Region von $\alpha 1(IV)$ zur Verfügung gestellt wurden. Danach wurde für $\alpha 1(IV)$ ein Oligonukleotid mit einer eindeutigen Sequenz hergestellt (siehe Teil I,2., S. 18). Somit war es nur für $\alpha 2(IV)$ erforderlich, ein Oligonukleotid nach den abgeleiteten DNA-Sequenzen der NC1-Region zu synthetisieren.

Bei der Auswahl eines geeigneten Sequenzabschnitts für die Oligonukleotidsynthese müssen folgende wichtige Kriterien berücksichtigt werden:

1. Es müssen Peptidabschnitte mit Aminosäuren gewählt werden, die nur durch wenige Codons bestimmt werden.
2. Die Oligonukleotide müssen eine Mindestlänge von 14 Basen haben.

Aufgrund der Degeneration des genetischen Codes erhält man bei der Ableitung von Nukleinsäuren von Peptiden immer mehrere mögliche Sequenzen. Um die Anzahl der verschiedenen Nukleotidkombinationen möglichst gering zu halten, müssen für die Oligonukleotidsynthese Sequenzabschnitte mit Aminosäuren ausgesucht werden, die nur durch ein oder zwei Codons bestimmt werden. Trotzdem kann in Abhängigkeit von der Länge der Sequenz die Anzahl der verschiedenen Kombinationen so groß werden, daß die Spezifität eines entsprechenden Oligonukleotidgemischs bei der Hybridisierung zu gering wäre. Die verschiedenen möglichen Sequenzen müssen dann in mehreren Teilgemischen synthetisiert und getrennt eingesetzt werden.

Eine andere Methode, die Anzahl der verschiedenen Sequenzen zu limitieren, ist die Analyse des Codongebrauchs bereits bekannter Gene. Der Ausschluß kaum benutzter Codons bei der Oligonukleotidsynthese ist allerdings mit dem Risiko verbunden, daß im Einzelfall die richtige Sequenz möglicherweise doch nicht enthalten ist.

Die Länge eines Oligonukleotids ist entscheidend für die Stabilität und die Spezifität der Hybridisierung. Die Oligonukleotide müssen daher eine Mindestlänge haben. Bei Einsatz von Oligonukleotiden in eukaryontischen Systemen darf wegen der Spezifität eine Länge von 14 Basen nicht unterschritten werden. Erst dann ist gewährleistet, daß die fragliche Sequenz in einer cDNA-Genbank bzw. mRNA-Population der Komplexität höherer Eukaryonten höchstens einmal vorkommt (Szostak et al., 1979). Ansonsten muß bei Oligonukleotidgemischen hinsichtlich der optimalen Länge ein Kompromiss eingegangen werden. Mit der Länge der Sequenzen steigen zwar Stabilität und Spezifität der Bindungen, gleichzeitig erhöht sich jedoch auch die Anzahl der verschiedenen Sequenzen im Gemisch.

Für den Einsatz der Oligonukleotide zur Gewinnung von Typ IV-spezifischer cDNA kamen zwei mögliche Strategien in Frage:

1. Verwendung der Oligonukleotide als Hybridisierungsproben für cDNA-Genbanken, deren cDNA-Synthese am Poly(A)-Schwanz der mRNA mit Oligo(dT) als Primer gestartet worden war;
2. Einsatz der Oligonukleotide als Primer zur Synthese von spezifischen cDNA-Genbanken.

Ausschlaggebend für die Art der Vorgehensweise ist der Abstand der verfügbaren Oligonukleotidsequenzen, bzw. des korrespondierenden Genabschnitts, zum 3'-Ende der mRNA. Bei der Hybridisierung von Genbanken, deren cDNA ausgehend vom Poly(A)-Schwanz am 3'-Ende der mRNA extendiert worden ist, darf die Distanz der fraglichen Nukleotidsequenzen zum Startpunkt der cDNA-Synthese nicht zu groß sein. Mit zunehmender Entfernung vom 3'-Ende sinkt nämlich aufgrund von Abbrüchen der cDNA-Synthese die Wahrscheinlichkeit, positive Klone zu identifizieren. Oligonukleotide sind somit nur dann mit Erfolg als Hybridisierungsprobe für derartige cDNA-Genbanken einzusetzen, wenn die fraglichen Sequenzen einen Abstand von zwei- bis dreitausend Basen zum 3'-Ende der mRNA nicht überschreiten. Ist dies nicht gewährleistet, dann empfiehlt sich ein Vorgehen nach der zweiten Methode, der Primerextension zur Synthese einer spezifischen cDNA-Genbank.

Die codierenden Sequenzen der NC1-Domäne vom Carboxyterminus des Proteins sind im 3'-Bereich der Gene bzw. der mRNA lokalisiert. Anhand der mRNA-Längen (siehe Teil I, 2., S. 21) konnte man davon ausgehen, daß die Distanz der NC1-Sequenzen zum 3'-Ende der mRNA 1.5 bis 2 kb beträgt. Bei Einsatz von Oligonukleotiden aus der NC1-Region als Hybridisierungsproben bestanden daher gute Aussichten, aus einer cDNA-Genbank, deren cDNA-Synthese an Oligo(dT) als Primer extendiert worden war, positive cDNA-Klone zu isolieren.

Ein Vorgehen nach dieser Methode wurde auch aus einem weiteren Grund vorgezogen. Allein durch den Einsatz der Oligonukleotide zur Primerextension wäre es nicht möglich gewesen, den gesamten NC1-Bereich zu erfassen, einschließlich eventuell vorhandener C-terminaler Propeptidsequenzen (siehe Einleitung). Hierzu hätten Oligonukleotide als Primer eingesetzt werden müssen, die an das 3'-Ende der codierenden Region der $\alpha 1$ - bzw. $\alpha 2$ -mRNA binden, da nur dann in der extendierten cDNA der vollständige 3'-Bereich der NC1-Sequenzen enthalten sein kann. Die C-terminalen Peptidsequenzen der NC1-Domäne waren jedoch nicht bekannt.

Entscheidend für eine erfolgreiche Isolierung von spezifischen cDNA-Klonen ist außerdem, daß für die cDNA-Synthese eine geeignete RNA zur Verfügung steht. Die verwendete RNA-Präparation sollte einen möglichst hohen Gehalt der spezifischen mRNA-Spezies enthalten. Hier sind molekularbiologische Untersuchungen von Basalmembrankomponenten wie Typ IV Kollagen mit einem Problem konfrontiert. Die Zellpopulationen, die in vivo für Basalmembranproteine codierende mRNA synthetisieren, sitzen als dünne Zellschicht auf den Basalmembranen auf und sind zusammen mit ihrem Substrat im Organismus verbreitet. Folglich gestaltet sich die Isolierung dieser Zellen zur RNA-Gewinnung äußerst schwierig. Zudem sind Basalmembranen sehr stabile Strukturen, die im adulten Organismus nur langsam (vierzig Tage bis zwei Jahre) erneuert werden (Walker, 1972). Die fraglichen Zellen dürften folglich nur in sehr geringem Umfang Basalmembranproteine bzw. die entsprechende mRNA synthetisieren.

Schon im Laufe der proteinchemischen Analysen von Basalmembranproteinen war man zunehmend auf bestimmte Tumore und Tumorzelllinien, die vermehrt Basalmembrankomponenten synthetisieren, als Materialquelle umgestiegen. Diese Tumorsysteme bilden auch genügend spezifische mRNA, so daß molekularbiologische

Arbeiten über Basalmembrankomponenten möglich sind. Zudem bieten sie den Vorteil, daß sie schnell wachsen und ohne Kontamination durch andere Zelltypen kultiviert werden können.

Im Rahmen der vorliegenden Arbeit wurden RNA-Präparationen bzw. cDNA aus folgenden Tumorsystemen verwendet. Der bereits in der Einleitung erwähnte EHS (Engelbreth-Holm-Swarm)-Maustumor (Orkin et al., 1977) hat als extensiver Basalmembranproduzent schon viel zur Charakterisierung von Basalmembranproteinen beigetragen und ist daher auch zur Gewinnung von Basalmembranspezifischer mRNA geeignet. Er kann bei Mäusen leicht transplantiert werden. Eine RNA-Präparation aus dem EHS-Tumor wurde großzügigerweise von M. Laurent (Paris) zur Verfügung gestellt.

Außerdem wurden RNA-Präparationen aus zwei Tumorzelllinien der Maus eingesetzt, und zwar der PYS-2- und der F9-Zelllinie (Lehmann et al., 1974; Strickland et al., 1980). Bei den PYS-2-Zellen (Parietal Yolk Sac) handelt es sich um differenzierte Teratokarzinomzellen, die in vieler Hinsicht embryonalen Endodermzellen ähneln. Sie exprimieren Basalmembranproteine wie Typ IV Kollagen und Laminin und scheiden sie ins umgebende Medium aus (Oberbäumer et al., 1982; Leivo et al., 1982).

Die F9-Zelllinie stammt ebenfalls aus einem Teratokarzinom der Maus. Im undifferenzierten Zustand produzieren diese Zellen keine Basalmembranproteine. Sie differenzieren jedoch in Gegenwart von Vitamin A-Säure und Dibutyryl-cAMP zu endodermartigen Zellen, die dann in hohem Umfang Basalmembrankomponenten synthetisieren (Strickland et al., 1980). Auch nicht induzierte F9-Zellen differenzieren bei konfluenter Kultur bis zu einem gewissen Grad spontan. Dies wird in den Ergebnissen von Northernhybridisierungen (siehe Teil III,4., S. 103) deutlich, wo auch bei nicht induzierten F9-Zellen Typ IV-spezifische mRNA nach-

zuweisen ist. Die hierbei verwendete RNA wurde von Y. Yamada (Bethesda, U.S.A.) zur Verfügung gestellt.

Ferner wurde humane RNA isoliert, und zwar aus der Zelllinie HT-1080 (Alitalo et al., 1980). Hierbei handelt es sich um etablierte Fibrosarkomzellen, die ebenfalls Basalmembranproteine synthetisieren.

2. Typ IV-spezifische Oligonukleotide und Analyse ihres Hybridisierungs- verhaltens in Northernhybridisierungen

Abb.5 zeigt schematisch die bekannten Peptidsequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$ bei Mensch und Maus, sowie die Positionen der Oligonukleotide, die zur Isolierung von Typ IV-spezifischen cDNA-Klonen eingesetzt wurden.

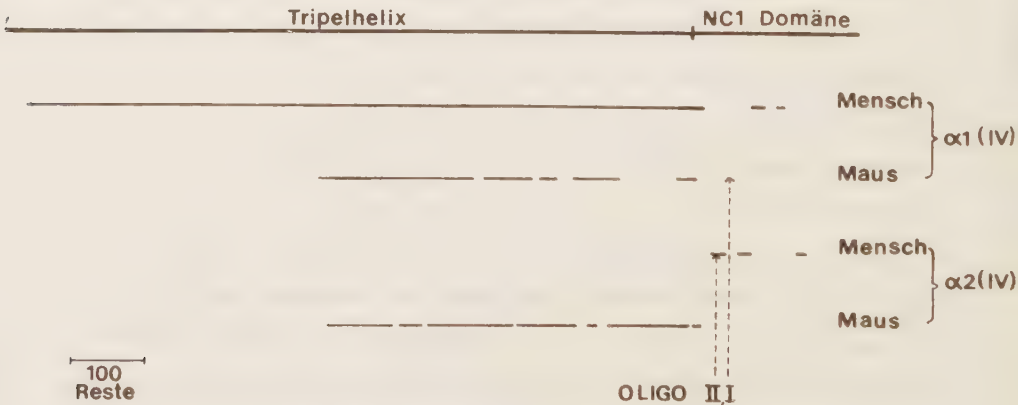


Abb.5: Durch Proteinsequenzierung bekannte Peptidsequenzen der beiden α -Ketten von Typ IV Kollagen von Mensch und Maus und Positionen der Oligonukleotide Oligo I ($\alpha 1$) und Oligo II ($\alpha 2$)

Bei $\alpha 1(IV)$ war es, wie bereits erwähnt, nicht erforderlich, Oligonukleotidsequenzen von den bekannten Peptidabschnitten abzuleiten. Hier war bereits ein kurzer cDNA-Abschnitt (150 bp) der NC1-Region von Maus aufgeklärt. Die entsprechenden Informationen wurden von M. Kurkinen (Rutgers, U.S.A.) zur Verfügung gestellt. Die Lokalisation der fraglichen Sequenzen innerhalb der NC1-Region von $\alpha 1(IV)$ (Aminosäurepositionen 26-75 in Abb.14, S. 49) war zu Beginn der Arbeiten noch nicht bekannt. Nach dieser cDNA wurde ein Oligonukleotid (Oligo I) synthetisiert. Oligo I (Tab.3) ist 20 Basen lang und komplementär zu dem Abschnitt, der für die Aminosäuren Pos. 36 bis 42 der NC1-Region von $\alpha 1(IV)$ der Maus codiert (siehe auch Abb.14).

Tab.3: $\alpha 1(IV)$ - und $\alpha 2(IV)$ -spezifische Oligonukleotide

Oligo I ($\alpha 1$)

Protein	(N-)	Val - Gln - Gly - Asn - Glu - Arg - Ala	(-C)
codierende DNA	(5')	GTC - CAA - GGC - AAC - GAG - CGT - GCC	(3')
Oligo I	(3')	CAG - GTT - CCG - TTG - CTC - GCA - CG	(5')

Oligo II ($\alpha 2$)

Protein	(N-)	Tyr - Phe - Glu - Gly - Gln - Glu - Lys	(-C)
codierende DNA	(5')	TAY - TTY - GAR - GGN - CAR - GAR - AAR	(3')
cDNA	(3')	ATR - AAR - CTY - CCN - GTY - CTY - TTY	(5')
Nuc 17		ATA - - - CCT - - - TT	(16x)
Nuc 18		ATG - - - CCT - - - TT	(16x)
Nuc 19		ATA - - - CCR - - - TT	(32x)
Nuc 20		ATG - - - CCR - - - TT	(32x)

N=A, T, G, C

R=A, G

Y=T, C (nach Cornish-Bowden, 1985)

Für $\alpha 2(IV)$ wurde von den teilweise bekannten Peptidsequenzen des NC1-Bereichs eine Oligonukleotidsequenz abgeleitet und synthetisiert (Oligo II, Abb.5). Da hier nur Proteindaten von Mensch vorlagen, konnte nur ein Oligonukleotid mit humanen Sequenzen hergestellt werden. Zu Beginn dieser Arbeit stand allerdings nur eine cDNA-Genbank der Maus zur Verfügung. Man konnte jedoch damit rechnen, daß eine Kreuzhybridisierung von humanen Oligonukleotidsequenzen mit den Maussequenzen der Genbank stattfinden würde. Es hatte sich nämlich gezeigt, daß bei $\alpha 1(IV)$ in den bei Mensch und Maus bekannten korrespondierenden Peptidsequenzen über 90% der Aminosäuren identisch sind (Schuppan et al., 1982; Babel und Glanville, 1984). Daher war auch für $\alpha 2(IV)$ eine ähnlich hochgradige Inter-spezieshomologie zu erwarten. Es muß zwar berücksichtigt werden, daß bei Mensch und Maus eventuell eine unterschiedliche Codonbenutzung für identische Aminosäuren vorliegt - dies stellt jedoch kein Problem dar, wenn in die Oligonukleotidsynthese alle möglichen Codons einbezogen werden. Aufgrund dieser Überlegungen erschien es erfolgversprechend, eine cDNA-Genbank der Maus mit einem Oligonukleotid humanen Ursprungs zu hybridisieren.

Oligo II (20 Basen) wurde von der Peptidsequenz Pos. 11 bis 17 des Bromcyanpeptids CB 3 (Siebold, Dissertation 1986) der NC1-Region der $\alpha 2$ -Kette von Mensch abgeleitet (Tab.3). Der betreffende Peptidabschnitt enthält nur Aminosäuren, die durch zwei Codons bestimmt werden, mit Ausnahme von Glycin, dem vier Codons zugeordnet sind. Danach ergeben sich 128 verschiedene mögliche DNA-Sequenzen. Diese hohe Anzahl wurde durch den Ausschluß des Codons GGG für Glycin auf 96 reduziert. Wurden diese 96 Varianten in einem einzigen Gemisch synthetisiert, dann hätte die richtige Sequenz einen Anteil von nur ca. 1 % in der Hybridisierungsprobe, die Spezifität wäre also sehr gering. Daher wurden die 96 Sequenzen zur Synthese in vier Teilgemische (Nuc 17 bis Nuc 20) aufgeteilt.

Nuc 17 und Nuc 18 enthalten jeweils 16 verschiedene Sequenzen, Nuc 19 und Nuc 20 jeweils 32.

Das Codon GGG für Glycin wurde bei der Synthese von Oligo II deshalb nicht berücksichtigt, weil dieses Codon in den bereits bekannten Kollagen-Genen verschiedener Organismen fast nicht vorkommt (Boedtker et al., 1983). Später zeigte sich allerdings, daß Typ IV Kollagen in dieser Hinsicht deutlich von den interstitiellen Kollagenen abweicht, da das Codon GGG für Glycin sogar relativ häufig auftritt (siehe Teil II, 3.2., S. 75). Glücklicherweise erwies sich im nachhinein, daß die fragliche Position bei $\alpha 2(IV)$ durch ein anderes Codon (GGC) bestimmt wird (Pos. 38 in Abb.14, S. 49).

Die entscheidende Frage war zunächst, ob Oligo II, das ja $\alpha 2(IV)$ -Humansequenzen enthält, mit $\alpha 2(IV)$ -Sequenzen von Maus kreuzhybridisieren würde. Wie bereits erwähnt, stand für den geplanten Einsatz der Oligonukleotide zur Isolierung von Typ IV-spezifischer cDNA nur eine cDNA-Genbank von Maus zur Verfügung. Außerdem wäre das "screening" einer Genbank mit mehreren Teilgemischen von Oligonukleotiden mit zu großem Aufwand verbunden, so daß eine Vorauswahl unter den Teilgemischen stattfinden muß. In Northernhybridisierungen läßt sich feststellen, ob die erwartete Hybridisierung auftritt, und welches der Gemische die richtige Sequenz enthält; Voraussetzung hierfür ist die Kenntnis der Länge der spezifischen mRNA. Für $\alpha 1(IV)$ und $\alpha 2(IV)$ war bekannt, in welchem Größenbereich die mRNAs auftreten. Für beide mRNAs waren Längen von mindestens 5.1 kb zu erwarten, da sich die Peptidketten aus ca. 1700 Aminosäuren zusammensetzen (Dean et al., 1983). Aus einer Arbeit von Kurkinen et al. (1983) war zu ersehen, daß die durch differentielle Hybridisierung isolierten $\alpha 1(IV)$ -Klone gegen mRNA im Bereich von 6.8 kb hybridisierten. Dies war ein Indiz dafür, daß zumindest die $\alpha 1$ -mRNA eine relativ lange nichttranslatierte Region enthält. Vor allem aber

war es ein Anhaltspunkt, in welcher Größenordnung die Typ IV Kollagen mRNAs anzusiedeln sind.

Neben der Identifizierung der richtigen Teilmischung ist es außerdem wichtig, anhand von Northernhybridisierungen die optimalen Hybridisierungsbedingungen festzustellen, um unspezifische Bindungen möglichst auszuschalten.

Eine große Rolle für die Spezifität der Hybridisierung von synthetischen Oligonukleotiden spielt die Hybridisierungstemperatur. Die Auswahl eines geeigneten Temperaturbereichs muß unter Berücksichtigung des Schmelzpunktes (T_m) der jeweiligen DNA-Probe - das ist die Temperatur, bei der 50% der DNA einzelsträngig vorliegen - erfolgen. Dieser Schmelzpunkt einer DNA-Probe hängt von verschiedenen Parametern ab, wie Basenlänge der Probe, GC-Gehalt, sowie von der Ionenstärke und dem pH-Wert der Hybridisierungslösung. Wenn DNA-Proben kürzer sind als 50 Basen, ist der GC-Gehalt der maßgebliche Faktor für die Stabilität der Hybridisierung. Der Einfluß von Ionenstärke und pH-Wert in der Hybridisierungslösung können dann unter Standardbedingungen (die Na^+ -Konzentration sollte über 0.4 M liegen, der pH-Wert zwischen 5 und 9) vernachlässigt werden (Meinkoth und Wahl, 1984).

Für Oligonukleotide bis zu 20 Basen Länge kann der Schmelzpunkt (T_m) durch die folgende, aus experimentellen Daten abgeleitete Beziehung, bestimmt werden:

$$T_m(^{\circ}C) = 4 (G+C) + 2 (A+T)$$

(Meinkoth und Wahl, 1984)

In diese Formel gehen die Länge und die Basenzusammensetzung der DNA-Probe ein, also die beiden für das Hybridisierungsverhalten von kurzen Oligonukleotiden entscheidenden Faktoren.

Bei der Hybridisierung lassen sich besten Ergebnisse erzielen, wenn man die Inkubation zunächst bei hoher Temperatur über dem T_m -Wert der Probe startet und dann langsam auf eine niedrigere Hybridisierungstemperatur übergeht. Es empfiehlt sich, die weitere Hybridisierung bei einer Temperatur durchzuführen die 10-20 °C unter dem T_m -Wert der Probe liegt, die Filter bei dieser Temperatur auch zu waschen, und dann ein oder mehrere Waschschrirte bei höherer Temperatur anzuschließen (Meinkoth und Wahl, 1984). Durch die anfangs hohe und dann abnehmende Stringenz der Hybridisierungstemperatur wird erreicht, daß sich stabile, spezifische Bindungen zuerst ausbilden. Die spezifischen Bindungsstellen werden also zuerst besetzt und sind damit beim Übergang zu niedrigeren Temperaturen weniger zugänglich für unspezifische Hybridisierungen. Bei Temperaturen unter dem Schmelzpunkt sind dann spezifische Hybridisierungen stabiler und entsprechend zeigen sich intensivere Markierungen, begünstigt werden allerdings auch unspezifische Bindungen mit Basenfehlpaarungen. Bei steigender Waschtemperatur sollte sich dann anhand von Veränderungen im Hybridisierungsmuster zwischen stabilen und spezifischen Bindungen und unspezifischen Bindungen unterscheiden lassen.

Bei Verwendung von Oligonukleotidgemischen ist die Wahl geeigneter Hybridisierungsbedingungen stark erschwert. Die einzelnen Sequenzen im Gemisch können in ihrer Basenzusammensetzung und damit hinsichtlich ihres jeweiligen Schmelzpunktes stark variieren. Dadurch kann eine GC-reiche Sequenz trotz Basenfehlpaarung eine stabilere Bindung ausbilden als eine weniger GC-reiche, aber dafür spezifische Sequenz. Auch sind die Auswirkungen von Basenfehlpaarungen auf die Stabilität der Hybride sehr unterschiedlich, abhängig davon, welche Basen betroffen sind, und auch von der Position der Fehlpaarung innerhalb der Sequenz. Basenfehlpaarungen können Schmelzpunkterniedrigungen von wenigen °C bis zu

12 °C bedingen (Wallace et al., 1979). Die nicht kalkulierbaren Einflüsse des GC-Gehalts auf das Hybridisierungsverhalten können durch Zugabe von Tetramethylammoniumchlorid zur Waschlösung eliminiert werden (Wood et al., 1985). Diese Substanz bindet selektiv an AT-Paare und erhöht die Stabilität der Bindung. Dadurch wird der Unterschied der Stabilitäten von AT- und GC-Paaren aufgehoben, der Schmelzpunkt einer Probe hängt dann nur noch von der Basenlänge ab. Diese Methode war zum Zeitpunkt der im folgenden beschriebenen Northernhybridisierung noch nicht bekannt.

Die Auswahl der richtigen Teilmischung sowie geeigneter Hybridisierungsbedingungen ist also bei einem geplanten Einsatz von Oligonukleotidmischungen sehr wichtig. Dies soll anhand der Northernhybridisierung mit der $\alpha 2(IV)$ -spezifischen Mischung Oligo II (Tab.3) ausführlich beschrieben werden. Im Anschluß daran wird nur kurz auf die Northernhybridisierung mit Oligo I eingegangen. Da es sich hier um eine eindeutige Sequenz handelte, war die Analyse und Interpretation des Hybridisierungsverhaltens weitaus einfacher als bei Oligo II.

Für den RNA-Transfer (siehe Methodenteil, S. 144) wurde Poly(A)+mRNA aus der Maus Zelllinie PYS-2 (siehe Einführung) verwendet. Die vier Teilgemische Nuc 17 bis Nuc 20 von Oligo II wurden durch Polynukleotidkinase radioaktiv (^{32}P - γ -ATP) endmarkiert und jeweils adäquate Mengen an Radioaktivität (5 Mill. cpm von Nuc 17 und Nuc 18, sowie 10 Mill. cpm von Nuc 19 und Nuc 20) mit einem Filter mit transferierter PYS-2 Poly(A)+RNA hybridisiert. Die Teilgemische Nuc 17 bis 20 (Tab.3.) variieren in ihrem GC-Gehalt zwischen 25 % und 55 %. Nach der oben aufgeführten Beziehung wurden Schmelzpunkte zwischen 50 °C und 62 °C errechnet. Unter Berücksichtigung der oben genannten Kriterien wurde daher bei 43 °C hybridisiert. Auch die erste Waschung der Filter wurde bei dieser Tem-

peratur durchgeführt. Eine 20-stündige Exponierung ergab die in Abb. 6 gezeigten Bilder.

Das Hauptinteresse galt zunächst der Frage, ob die erwartete Kreuzhybridisierung von Oligo II mit der $\alpha 2(IV)$ -mRNA von Maus stattgefunden hatte. Aus den Resultaten (Abb.6) ging hervor, daß dies offensichtlich der Fall war. Nuc 19 hybridisierte an eine Bande im Bereich von 6 kb, die in der für die $\alpha 2$ -mRNA zu erwartenden Größenordnung lag, daneben aber auch sehr stark mit der 28S rRNA. Nuc 20 hybridisierte - bei starkem "background" - ebenfalls an eine RNA-Bande (6.5 kb), die im fraglichen Bereich lag. Bei Nuc 17 und Nuc 18 war keine Bindung an eine definierte RNA-Bande sichtbar. Dies änderte sich auch nach einer Exponierungszeit von mehreren Tagen nicht. Die richtige $\alpha 2(IV)$ -Sequenz mußte also entweder in Nuc 19 oder in Nuc 20 enthalten sein.

Um zwischen spezifischen und unspezifischen Bindungen von Nuc 19 und Nuc 20 besser unterscheiden zu können, wurden die beiden Filter ein zweitesmal unter stringenteren Temperaturbedingungen bei 53 °C (6xSSC) gewaschen. Danach war eine Exponierung von 5 Tagen notwendig, um eine vergleichbare Intensität der Banden mit der nur 20-stündigen Exponierung nach dem Waschen bei 43 °C zu erhalten (Abb.6). Bei Nuc 19 zeigte sich überraschenderweise ein stark verändertes Bandenmuster. Die Kreuzhybridisierung mit der 28S rRNA, sowie die bei 43 °C-Waschtemperatur noch deutlich sichtbare Markierung der 6 kb Bande waren kaum bzw. gar nicht mehr festzustellen. Stattdessen war eine 6.5 kb Bande dominant. Offensichtlich hatte die kurze Exponierungszeit (20 h) nach dem Waschen bei 43 °C nicht ausgereicht, diese Bande hervortreten zu lassen. Bei dem Gemisch Nuc 20 änderte sich das Muster durch die Erhöhung der Waschtemperatur kaum. Neben hohem Hintergrund war nachwievor eine Bande bei 6.5 kb erkennbar.

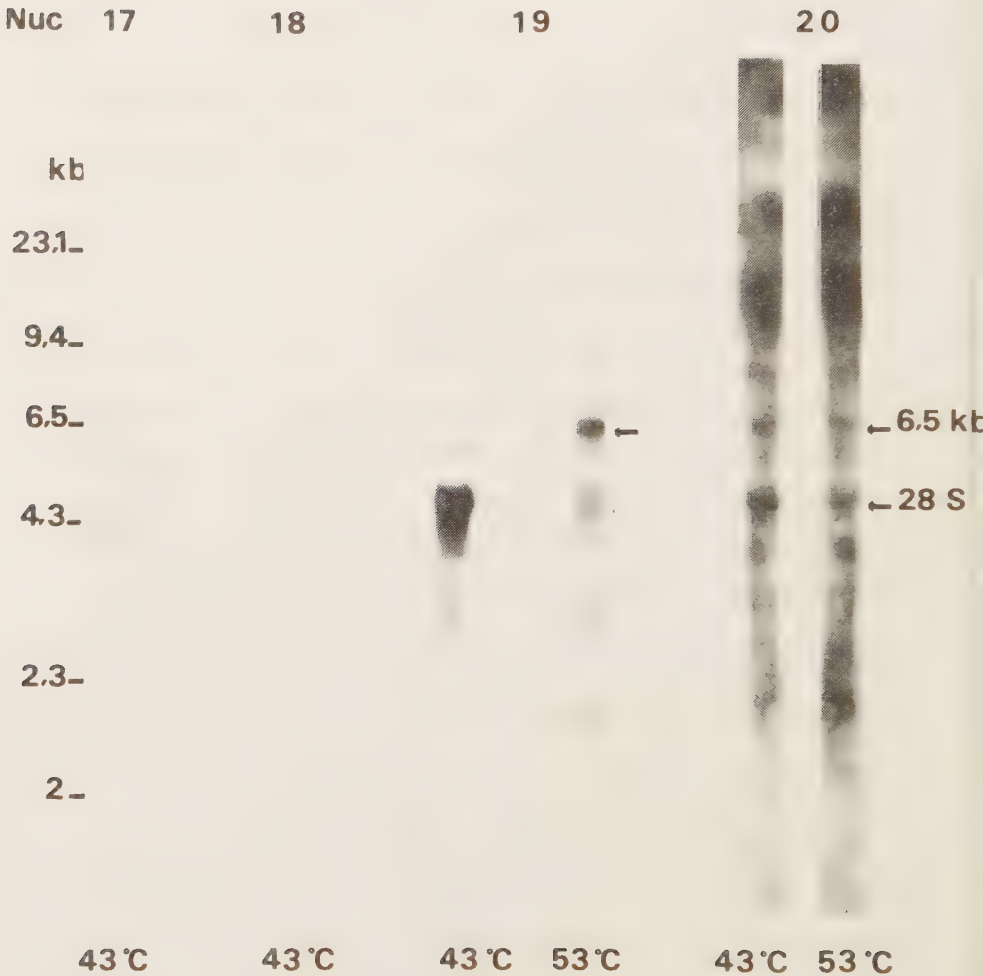


Abb. 6: Northernhybridisierungen mit den Gemischen Nuc 17 bis Nuc 20 des $\alpha 2(\text{IV})$ -spezifischen Oligonukleotids Oligo II
ca. 10 μg von PYS-2 Poly(A)+RNA wurden durch Glyoxal denaturiert, in einem 1%igen Agarosegel aufgetrennt und durch einen Elektroblot auf Nitrozellulose transferiert. Die Hybridisierungsbedingungen sind im Text beschrieben.

Als Großenmarker wurde Hind III-verdaute λ -DNA eingesetzt. Eine Bande von 6.5 kb und die 28S rRNA sind durch Pfeile gekennzeichnet.

Exponierungszeiten: 20 h nach dem Waschen bei 43 °C
120 h nach dem Waschen bei 53 °C

Aus diesen Ergebnissen wurde abgeleitet, daß Nuc 19 die richtige $\alpha 2(IV)$ -Sequenz enthält. Die durch Nuc 19 markierte Bande von 6.5 kb liegt im erwarteten Größenbereich und zeigt zudem unter stringenten Hybridisierungsbedingungen die stärksten Signale. Die bei 43 °C-Waschtemperatur beobachteten Bindungen von Nuc 19 an die 6 kb und die 28S RNA dürften das Resultat von unspezifischer Hybridisierung mit Fehlpaarungen gewesen sein. Darauf deutet die Tatsache hin, daß diese Bindungen bei einer Waschtemperatur von 53 °C nicht mehr stabil sind. Ähnliches gilt vermutlich für das Hybridisierungsmuster von Nuc 20. Nuc 20 bindet zwar auch an die fragliche 6.5 kb Bande, jedoch ist die Intensität nach dem Waschen bei 53 °C deutlich reduziert in Anbetracht der weitaus längeren Exponierungszeit. Dieser beobachtete Effekt dürfte ebenfalls die Folge von Fehlpaarungen im Hybrid sein, die bei 43 °C noch nicht die Effizienz der Bindung von Nuc 20 an die RNA herabsetzen, sondern erst bei einer höheren Waschtemperatur.

Das auf diese Weise selektionierte Gemisch Nuc 19 der $\alpha 2(IV)$ -spezifischen Mischung Oligo II wurde somit als Hybridisierungsprobe für eine cDNA-Genbank eingesetzt. Die Hybridisierung mußte bei einer Temperatur von 53 °C stattfinden, da aufgrund der Northernhybridisierung bei niedrigerer Temperatur eine starke unspezifische Hybridisierung vor allem mit ribosomalen Sequenzen zu erwarten war. Man mußte nämlich davon ausgehen, daß in der eingesetzten cDNA-Genbank auch zahlreiche rRNA-Klone vertreten sind, die durch "selfpriming" entstanden sind (G. Vogeli, persönliche Mitteilung).

Nach Isolierung und Sequenzierung eines $\alpha 2(IV)$ -spezifischen cDNA-Klones (siehe Teil I,4., S. 38) zeigte sich, daß Nuc 19 tatsächlich die authentische $\alpha 2$ -Sequenz von Maus enthält.

Obwohl es sich bei dem $\alpha 1(\text{IV})$ -spezifischen Oligonukleotid Oligo I um eine eindeutige Sequenz handelte, wurde das Hybridisierungsverhalten ebenfalls in einer Northernhybridisierung analysiert. Auch hier wurde Poly(A)+RNA aus der PYS-2-Zelllinie (Maus) eingesetzt. Unter stringenten Temperaturbedingungen bei 50 °C hybridisierte Oligo I an zwei RNA-Banden mit einer Länge von 6.7 kb und 5.4 kb (nicht gezeigt), wobei die 6.7 kb-Bande intensiver war. Die Markierung einer 6.7 kb-mRNA stimmte gut überein mit den Ergebnissen von Kurkinen et al. (1983), die eine Länge von 6.8 kb angegeben haben. Allerdings stellte sich die Frage, was dem Auftreten der 5.4 kb-Bande zugrunde liegt. Kreuzhybridisierung mit der $\alpha 2(\text{IV})$ -mRNA konnte weitgehend ausgeschlossen werden, nachdem Oligo II keine mRNA im Bereich von 5.4 kb markierte (Abb.6). Daher lag die Vermutung nahe, daß es sich bei der 5.4 kb-Bande um spezifische Markierung einer kürzeren Variante der $\alpha 1(\text{IV})$ -mRNA handelte. Bei fast allen bisher analysierten Kollagenen ist nämlich eine Längenheterogenität der mRNA ausgeprägt, wie später näher erläutert wird (Teil III), und auch bei $\alpha 1(\text{IV})$ zeigte sich, daß dies der Fall ist.

Nach Isolierung von zwei cDNA-Klonen (p5 und p9, Teil I,3., S. 29) durch Einsatz von Oligo I und deren Sequenzierung stellte sich heraus, daß die cDNA an einer Position eine andere Base aufweist als das Oligonukleotid. Für die Aminosäure Valin in Position 36 (Abb.14, S. 49) wurde bei beiden Klonen das Codon GTT identifiziert, von M. Kurkinen dagegen ein GTC (Tab.3). Für derartige Diskrepanzen kommen neben Sequenzierungsfehlern auch andere Ursachen in Frage, die an anderer Stelle (Teil II,2.1., S. 61) diskutiert werden. Jedenfalls konnte Oligo I erfolgreich zur Isolierung von Klonen eingesetzt werden - die Basenfehlpaarung hat die Spezifität von Oligo I offensichtlich nicht entscheidend beeinträchtigt, vermutlich, weil sie an einem Ende der Sequenz lokalisiert ist.

3. Isolierung $\alpha 1$ (IV)-spezifischer cDNA-Klone

3.1. Identifizierung und Charakterisierung der Klone p5 und p9

Oligo I (siehe Teil I,2.) wurde mit einer cDNA-Genbank hybridisiert, um Klone mit Sequenzen der NC1-Domäne der $\alpha 1$ -Kette von Typ IV Kollagen zu isolieren. Die verwendete cDNA-Genbank wurde freundlicherweise von G. Vogeli und M. Laurent zur Verfügung gestellt (Vogeli et al., 1985). Die für die Konstruktion der cDNA-Genbank verwendete RNA stammte aus dem EHS-Maustumor (Orkin et al., 1977). Poly(A)+RNA war über einen Sucrose-Gradienten nach Größe fraktioniert worden. Für die cDNA-Synthese wurden Fraktionen mit RNA größer als die 18S rRNA eingesetzt. Die cDNA war über GC-Schwänze in die Pst I-Stelle des Vektors pBR322 inseriert worden (Land et al., 1981). Für die Transformation war der E. coli-Stamm C600.4 (C600: Appleyard, 1954) verwendet worden.

Ein Aliquot der cDNA-Genbank (ca. 100 000 Transformanten) wurde auf 30 Nitrozellulosefiltern (82 mm Durchmesser) ausplattiert, angezüchtet und für die Hybridisierung vorbereitet (siehe Methodenteil, S. 147).

Für die Hybridisierung der Genbank (50 °C/ 5xSSC) wurden 16 Mill. cpm des radioaktiv markierten Oligo I eingesetzt (siehe Methodenteil, S. 145). Aus den ca. fünfzig positiven Signalen konnten nach zweimaliger Nachreinigung schließlich zwanzig Klone gewonnen werden. Zur leichteren Bestimmung der Insertgrößen wurden die isolierten Plasmide von zehn Klonen mit dem Restriktionsenzym Pst I inkubiert und so die integrierte cDNA herausgeschnitten. Hier zeigte sich, daß die Insertgrößen der Klone zwischen 300 bp und 600 bp lagen.

Die beiden längsten Klone (p5 und p9) wurden durch eine Restriktionskartierung weiter charakterisiert. Durch Spaltungen mit den Enzymen Pst I, Nco I und Pvu II ergaben sich die in Abb.7 gezeigten Bandenmuster. Bei Klon p9 war das Insert

bereits in den Vektor pUC9 (Vieira und Messing, 1982) kloniert worden, wie später erläutert wird; die cDNA von p5 war noch im Vektor pBR322 inseriert.

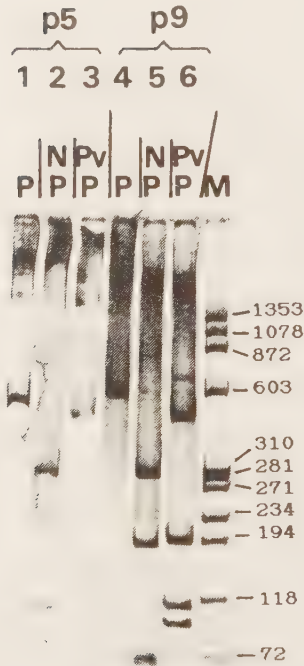


Abb. 7: Einfach- und Doppelspaltungen der Klone p5 und p9 mit Pst I, Nco I und Pvu II. Es wurden 0.2-0.4 µg Plasmid-DNA verdaut und anschließend auf einem 5-%igen Polyacrylamid-gel aufgetrennt. Die DNA-Banden wurden dann durch Silberfärbung (nach Merril et al., 1981) sichtbar gemacht. Klon p5 ist im Vektor pBR322 integriert, p9 im Vektor pUC9. Als Marker (M) wurde Hae III-geschnittene ϕ x-DNA eingesetzt.

P: Pst I
N: Nco I
Pv: Pvu II

Durch Spaltung mit Pst I (Spur 1 und 4) wurde jeweils das Insert herausgeschnitten, das bei beiden Klonen ca. 600 bp lang war. Bei p5 war eine interne Pst I-Schnittstelle vorhanden, deren Spaltung zu zwei Insertfragmenten von 540 bp und 75 bp führte.

Doppelverdau mit Pst I und Nco I zeigte, daß das 540 bp lange Pst I-Fragment von p5 durch Nco I in drei Subfragmente (320, 110 und 80 bp) zerlegt wurde

(Spur 2). Auch im Insert von p9 waren mindestens zwei Nco I-Stellen vorhanden, da drei Banden (310, 180 und 75 bp) sichtbar waren (Spur 5). Nachdem die kürzeste Bande von 75 bp ungewöhnlich breit und intensiv war, war zu vermuten, daß es sich hier um eine Doppelbande handelte und folglich noch eine dritte Nco I-Stelle in p9 vorhanden war. Im nachhinein bestätigte sich diese Vermutung - am 5'-Ende von p9 war eine dritte Nco I-Stelle lokalisiert, deren Spaltung ebenfalls zu einem 75 bp-Fragment führte (siehe Restriktionskarte in Abb.8).

Nach Doppelverdau mit Pst I und Pvu II zeigte sich, daß in beiden Klonen ein Pvu II-Stelle lokalisiert war. Bei p5 wurde das große Pst I-Fragment durch Spaltung mit Pvu II von 540 auf 500 bp reduziert (Spur 3). Das abgespaltene 40 bp Fragment war wegen seiner Kürze nicht mehr anfärbbar. Bei p9 wurde das Insert durch Pvu II in zwei Subfragmente von 500 und 100 bp unterteilt (Spur 6). Bei den zwei zusätzlich auftretenden Banden von 200 und 120 bp handelte es sich um Fragmente des Vektors pUC9, der zwei Pvu II-Spaltstellen besitzt.

Aus diesen Resultaten ging hervor, daß bei beiden Klonen jeweils eine Pvu II- und zwei Nco I-Stellen im Insert lokalisiert waren. Trotzdem war es nicht möglich, anhand der Restriktionsmuster in Abb.7 p5 und p9 zuzuordnen und den Überlappungsbereich zu identifizieren. Dies wurde erst nach einer Southernhybridisierung mit Oligo I erreicht. Da beide Klone durch Hybridisierung mit Oligo I isoliert worden waren, mußte die Sequenz des Oligonukleotids im Überlappungsbereich von p5 und p9 lokalisiert sein. Es wurden daher Einfach- und Doppelspaltungen von p5 und p9 mit Nco I und Pvu II durchgeführt, und durch eine Southernhybridisierung festgestellt, welche Fragmente durch Oligo I markiert werden. Anhand dieser Ergebnisse (nicht gezeigt) war es schließlich möglich, die Inserts von p5 und p9 zuzuordnen und eine Restriktionskarte aufzustellen (Abb.8). Es stellte sich heraus, daß die Überlappingsregion im Bereich einer gemeinsamen Pvu II- und Nco I-Stelle am 3'-Ende von p5 bzw. 5'-

Ende von p9 lag und nur ca. 80 bp umfaßte. Erfreulicherweise überdeckten p5 und p9 gemeinsam einen langen Sequenzabschnitt von über 1000 bp. Die Orientierung beider Klone in Bezug zum 3'- und 5'-Ende der mRNA konnte erst durch Sequenzierung festgestellt werden.

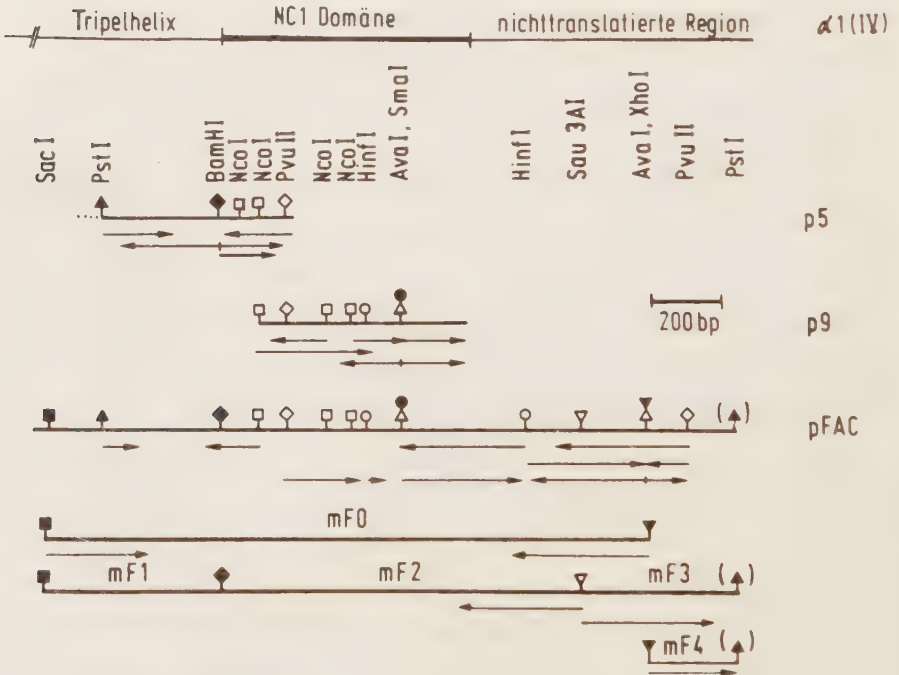


Abb. 8: Restriktionskarte der Klone p5, p9 und pFAC und Sequenzierungsstrategie
Richtung und Länge der durchgeführten Sequenzierungen sind durch Pfeile angezeigt. mF0 bis mF4 sind Subklone von pFAC in den Vektoren M13tg130 und tg131. Die Pst I-Stelle des Vektors pBR322 (Integrationsort der cDNA), die auch für Subklonierungen genutzt wurde, ist durch Klammern gekennzeichnet.

Zur Vereinfachung der geplanten Sequenzierung der beiden Klone p5 und p9 wurden die Inserts mit Pst I aus dem Vektor pBR322 herausgeschnitten und in die Pst I-Stelle der Polylinkerregion (Tab.4) des Vektors pUC9 (Viera und Messing, 1982) eingebaut.

Tab.4: Multiple Restriktionsstellen in den Polylinkerregionen des Plasmids pUC9 und der M13-Vektoren tg130 und tg131

pUC9

ACG GCA AGC TTG GCT GCA GGT CGA CCG ATC CCC GGG AAT TCA CTG GCC

HindIII Pst I Sal I Bam HI Sma I Eco RI Hae III
 Acc I Xba I
 Hind II

M13tg130

ACG AAT TCC CGG GAG AGC TCG ATA TCG CAT GCG GTA CCT CTA GAA GAA GCT TGG GAT CCG TCG ACC TGC AGC

[illegible]

W13tg131

ACG CAG ATC TGC AGG TCG ACG GAT CCC AAG CTT CTT CTA GAG GTA CCG CAT GCG ATA TCG AGC TCT CCC GGG AAT TCA

Bgl II Pst I Sal I Bam HI Hind III Xba I Kpn I Sph I Eco RV Sst I Sma I Eco RI
 Acc I Xba I
 Hind II

Der Hauptvorteil der pUC-Vektoren ist die Verfügbarkeit der zahlreichen singulären Restriktionsstellen in der Polylinkerregion, die die Klonierung von DNA-Fragmenten mit den verschiedensten Enden gestatten. Stark vereinfacht wird außerdem die chemische Sequenzierung nach Maxam und Gilbert (1980). Das integrierte Insert kann unter Ausnutzung der zahlreichen Restriktionsstellen direkt sequenziert werden, wie anschließend beschrieben wird (Abb.9), und damit

entfällt teilweise die aufwendige Isolierung von Insertfragmenten (Ruther et al., 1981).

Die Sequenzierung der Klone p5 und p9 erfolgte nach der chemischen Methode (Maxam und Gilbert, 1980). Am Beispiel der Analyse von Klon p5 soll das Vorgehen im einzelnen erläutert werden. Zur Sequenzierung vom 3'-Ende der integrierten cDNA aus wurde zunächst mit Sal I inkubiert (Abb.9), was eine Linearisierung des Plasmids zur Folge hat. Anschließend wurden die beiden 5'-

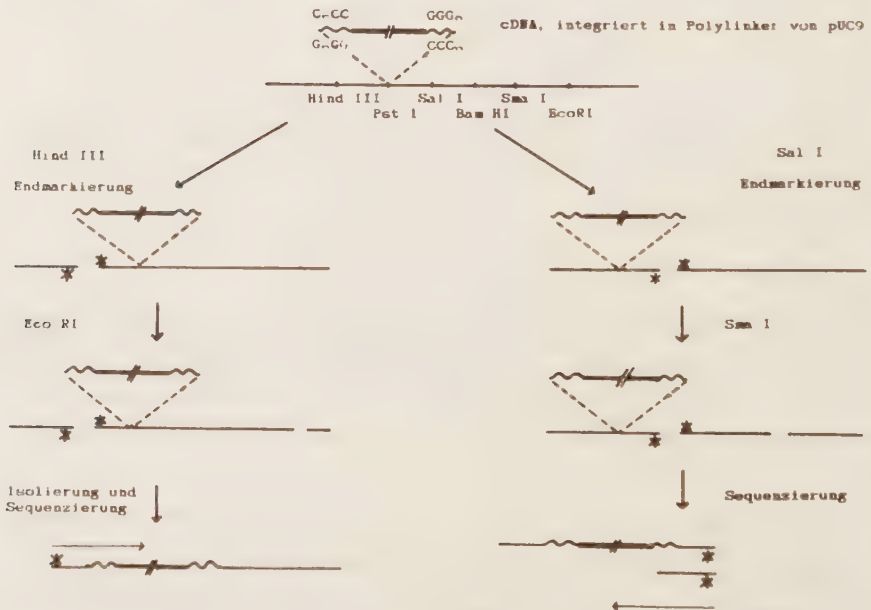


Abb.9: Sequenzierung von Klon p5 nach Maxam und Gilbert (1980) unter Ausnutzung der multiplen Restriktionsstellen in der Polylinker-region des pUC9

Die Sternchen kennzeichnen die radioaktiv (^{32}P) markierten 5'-Enden der DNA.

Die Sequenzierung vom 3'-Ende des Inserts aus (rechter Zweig) konnte direkt am Plasmid erfolgen. Für die Sequenzierung vom 5'-Ende aus (linker Zweig) mußte das Insert vom Vektor abgespalten und isoliert werden.

Enden mit γ -³²P-ATP durch Polynukleotidkinase radioaktiv markiert. Um das Insert sequenzieren zu können, mußte die Markierung auf der Vektorseite entfernt werden. Dies erfolgte durch einen Schnitt in der Sma I-Stelle der Polylinker-region. Das abgespaltene kurze Sal I/Sma I-Stück stört die anschließenden Sequenzierungsreaktionen nicht, seine Gegenwart verhindert lediglich die Identifizierung der ersten Basen. In diesem Fall handelt es sich um neun Basen, die sowieso vom Linker bzw. den dC-Schwänzen stammen, so daß von der eigentlichen cDNA-Sequenz nichts verloren geht. Für die Sequenzierung ausgehend vom 3'-Ende der cDNA mußte das Insert also nicht isoliert werden, sondern konnte direkt am Plasmid sequenziert werden.

Für die Sequenzierung vom 5'-Ende der cDNA aus wurde das Plasmid mit Hind III inkubiert (Abb.9) und ebenfalls radioaktiv markiert. Hier war es nicht möglich, die Markierung an der Vektorseite zu entfernen, da dafür keine weiteren Restriktionsstellen 5' von Hind III zur Verfügung standen. Daher mußte das Insert durch einen Schnitt mit Eco RI vom Vektor gelöst und über ein Polyacrylamidgel abgetrennt werden, um eine Sequenzierung zu ermöglichen.

Nachdem die cDNA von p5 soweit als möglich von den Flanken aus sequenziert worden war (Abb.8), mußten zur Ergänzung bzw. Absicherung der Sequenzdaten weitere Analysen von einer internen Schnittstelle aus erfolgen. Daher wurde ausgehend von der Bam HI-Schnittstelle im Insert (Abb.8) in beide Richtungen sequenziert. Hierzu mußten wiederum die entsprechenden Insertfragmente (zwei Bam HI/Pst I-Fragmente) aus Gelen isoliert werden.

Die Sequenzierungsstrategie für p5 wie auch für Klon p9 ist in Abbildung 8 zusammengefaßt. Es zeigte sich, daß p5 und p9 gemeinsam 654 Basen aus der NC1-Region enthalten, sowie 350 bp aus dem C-terminalen Bereich der Tripelhelix. Das fehlende Ende der NC1-Sequenz, nur 33 Basen, wie sich herausstellte, konnte schließlich mit Hilfe des cDNA-Klons pFAC bestimmt werden.

3.2. Charakterisierung von Klon pFAC

Dieser Klon wurde von Y. Yamada (Bethesda, U.S.A.) isoliert und im Rahmen der vorliegenden Arbeit sequenziert. Klon pFAC stammte aus einer Maus-cDNA-Genbank, die an Poly(A)+RNA aus induzierten F9-Zellen synthetisiert worden war, und war durch Hybridisierung mit einem Oligonukleotid (nach Pos. 66-70 der NC1-Domäne, Abb.14, S. 49) identifiziert worden (Oberbäumer et al., 1985). Auch hier war die cDNA über GC-Schwänze in den Vektor pBR322 inseriert worden.

Zunächst wurde durch eine Restriktionskartierung die Lage des ca. 2 kb langen pFAC in Bezug zu den Klonen p9 und p5 festgestellt. Es zeigte sich, daß pFAC die Sequenzen von p5 und p9 enthält, sowie am 5'-Ende weitere ca. 150 bp tripelhelikaler Sequenz und am 3'-Ende ca. 750 bp der nichtcodierenden Region (Abb.8).

Das Hauptinteresse galt zunächst der Vervollständigung der NC1-Sequenzen von $\alpha 1(IV)$. Zu diesem Zweck wurde ein Ava I/Hinf I-Fragment (Abb.8), das den fraglichen Bereich enthielt, in beiden Richtungen nach der chemischen Methode von Maxam und Gilbert sequenziert. Dadurch konnte das Ende der codierenden Region der NC1-Domäne, das durch zwei Stopcodons in kurzer Folge angezeigt wird (siehe Abb.25, S. 96) identifiziert werden. Durch die Analyse des komplementären Strangs dieses Ava I/Hinf I-Fragments wurden außerdem die Sequenzdaten der betreffenden Region abgesichert, nachdem bei p9 hier nur der codierende Strang sequenziert worden war (Abb.8). Auch in anderen Abschnitten war es erforderlich, wegen noch vorhandener Unklarheiten nach der Analyse von p5 und p9 die korrespondierenden Fragmente von pFAC zu sequenzieren, z. B. das Pvu II/Hinf I-Fragment innerhalb der NC1-Region. Schließlich wurde ein Teil (560 bp) der in diesem Klon enthaltenen nichttranslatierten Region am 3'-Ende nach der chemischen Methode (Maxam und Gilbert, 1980) sequenziert (Abb.8).

Um die noch unbekannten Bereiche am 3'- und 5'-Ende von Klon pFAC nach der im Vergleich zur chemischen Methode weniger aufwendigen Didesoxy-Sequenzierung nach Sanger (1980) aufzuklären, wurden M13-Subklone konstruiert. Das M13-System bietet, wie die pUC-Plasmide, die Vorteile der Verfügbarkeit multipler Restriktionsstellen in der Polylinkerregion. Nach Klonierungen in den M13-Vektor läßt sich die einzelsträngige Form des Phagen dann schnell isolieren und direkt für die Didesoxy-Sequenzierung einsetzen.

Die M13-Subklone wurden außerdem mit der Absicht konstruiert, für Northern-hybridisierungen und für spätere Hybridisierungen genomischer DNA einzelsträngige, leicht zu markierende Hybridisierungsproben zur Verfügung zu haben.

Insertfragmente von pFAC wurden in die Polylinkerregionen (Tab.4) der M13-Vektoren tg130 und tg131 (Kieny et al., 1983) eingesetzt (mF0 bis mF4 in Abb.8). Durch Sequenzierung des Subklons mF0 vom 5'-Ende aus wurden weitere 145 bp aus der Tripelhelix aufgeklärt, die in Klon p5 nicht mehr enthalten waren. Mit Hilfe der Subklone mF3 und mF4 wurde die Sequenz des 3'-nichttranslatierten Bereichs in Klon pFAC vervollständigt.

Durch Sequenzierung der Klone p5, p9 und pFAC wurde somit die Sequenz eines fast 2 kb langen Abschnitts $\alpha 1(IV)$ -spezifischer cDNA bestimmt. Dieser Bereich umfaßt die gesamte cDNA-Sequenz der NC1-Domäne (687 bp), sowie 495 bp der Tripelhelix und ca. 750 bp der nichttranslatierten Sequenz am 3'-Ende.

Bei allen beschriebenen Klonen fehlten die 3'-terminalen Sequenzen der RNA mit Anteilen des Poly(A)-Schwanzes. Diese Tatsache war überraschend, da die cDNA-Synthese am Poly(A)-Schwanz der mRNA mit Oligo(dT) als Primer initiiert worden war. Für dieses Phänomen gibt es zwei Erklärungen. Die verschiedenen cDNAs konnten durch Selbstprimen der mRNA entstanden sein. Fraglich ist jedoch, ob

die NC1-Region und die nichttranslatierte Region in dem Ausmaß stabile Sekundärstrukturen ausbilden können. Dies wäre eher für den tripelhelikalen Bereich mit seinen GC-reichen repetitiven Einheiten, die durch die periodisch auftretenden Glycin-Codons (GGN) und häufigen Prolin-Codons (CCN) zustande kommen, zu erwarten. Die wahrscheinlichere Erklärung für das Fehlen der 3'-Sequenzen ist, daß bei der Konstruktion der Genbanken die Synthese des zweiten Stranges der cDNA mit DNA-Polymerase vorzeitig abgebrochen ist. Dadurch wäre der 3'-Bereich der cDNA nur einzelsträngig vorgelegen und bei der anschließenden S1-Nukleasebehandlung abgebaut worden. Die Folge wäre gewesen, daß in der cDNA die 3'-terminalen Sequenzen nicht enthalten sind.

4. Identifizierung $\alpha 2$ (IV)-spezifischer cDNA-Klone

4.1. Isolierung und Charakterisierung der cDNA-Klone pAIIa und pAIIb

In Teil I,2. wurde gezeigt, daß Nuc 19, eines der vier Gemische des $\alpha 2$ (IV)-spezifischen Oligonukleotids Oligo II, in der Northernhybridisierung an eine RNA im Bereich von 6.5 kb bindet (Abb.6, S. 26). Die Bindung ist noch bei einer Inkubationstemperatur von 53 °C stabil, im Gegensatz zu den unspezifischen Hybridisierungen von Nuc 19 an eine 6 kb-Bande und an ribosomale RNA. Dieses Resultat wurde dahingehend interpretiert, daß in Nuc 19 die authentische $\alpha 2$ (IV)-Sequenz von Maus enthalten ist, und daß es sich bei der 6.5 kb Bande um die $\alpha 2$ (IV)-mRNA handelt.

Nuc 19 wurde daher mit Polynukleotidkinase radioaktiv (32 P) endmarkiert und gegen die EHS-cDNA-Genbank (Vektor: pBR322), aus der auch die $\alpha 1$ (IV)-Klone p5 und p9 stammten (Teil I,3.1.), hybridisiert. Die Ausbeute an positiven Signalen war sehr gering, nach zweimaliger Nachreinigung verblieben nur noch zwei Klone. Möglicherweise ist die beschränkte Anzahl an $\alpha 2$ -Klonen in der Genbank im

Vergleich zu den zahlreicher auftretenden $\alpha 1$ -Klonen darauf zurückzuführen, daß die Konzentration der $\alpha 2(IV)$ -mRNA in der Gesamt-RNA kleiner ist als die der $\alpha 1(IV)$ -mRNA. Jedenfalls ist auf Proteinebene ein Mengenverhältnis der beiden α -Peptidketten von ca. 2:1 ($\alpha 1:\alpha 2$) zu beobachten (Bächinger et al., 1982; Oberbäumer et al., 1982). In welchem Verhältnis die entsprechenden mRNAs vorliegen, ist noch nicht geklärt (siehe Diskussion).

Die wichtigste Frage war zunächst, ob das Oligonukleotidgemisch Nuc 19 spezifisch hybridisiert hatte, ob also die isolierten positiven Klone $\alpha 2(IV)$ -spezifische cDNA enthielten. Von großem Interesse war daher, dies möglichst schnell durch Sequenzierung der cDNA festzustellen. Erste Größenanalysen der beiden positiven Klone ergaben, daß das Insert des einen Klon nur wenige hundert Basenpaare umfaßte, während das des anderen ca. 1600 Basenpaare lang war. Der kurze Klon wurde nicht weiter charakterisiert. Der längere Klon, als pAlfa bezeichnet, wurde im Hinblick auf die geplante Subklonierung in M13-Vektoren zur Didesoxy-Sequenzierung nach Restriktionsstellen abgesucht, und zwar von Enzymen aus der Polylinkerregion (Tab.4) der Vektoren M13tg130 und tg131 (Kieny et al., 1983). Beim Verdau mit den Enzymen Bam HI, Eco RI, Hind III, Pst I, Sac I, Sal I und Sma I entstanden die in Abbildung 10 gezeigten Spaltmuster. Um die Anordnung der Insertfragmente zu erleichtern, war jeweils mit Pst I nachverdaut worden. Dadurch wurden sämtliche Insertanteile aus dem Plasmid herausgeschnitten, da ja die cDNA in die Pst I-Stelle des pBR322 inseriert worden war.

Es zeigte sich, daß für die Enzyme Sal I, Sma I und Hind III (Spur 9, 11 und 13) keine Erkennungsstellen im Insert von pAlfa vorliegen. Bam HI (Spur 4) und Sac I (Spur 7) schneiden jeweils einmal im Insert, Eco RI (Spur 2) und Pst I jeweils zweimal (Bam HI und Eco RI schneiden außerdem noch je einmal im pBR322). Die Aufstellung einer Restriktionskarte (Abb.11) zeigte, daß Doppel-

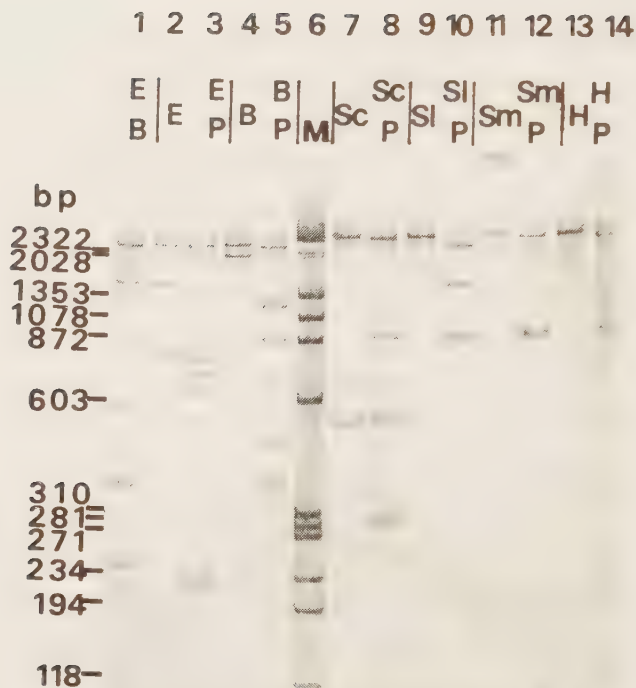


Abb. 10: Einfach- und Doppelspaltungen von Klon pAlIa mit verschiedenen Enzymen

Es wurden jeweils ca. 0.2 µg Plasmid-DNA verdaut und anschließend auf einem 5%igen Polyacrylamidgel aufgetrennt. Nach der Elektrophorese wurden die DNA-Banden durch eine Silberfärbung (nach Merril et al., 1981) sichtbar gemacht.

B: Bam HI

E: Eco RI

H: Hind III

P: Pst I

Sc: Sac I

Sl: Sal I

Sm: Sma I

M: DNA-Marker (λ/Hind III und φx/Hae III)

spaltung mit Eco RI und Pst I (Spur 3) die cDNA in fünf Stücke von ca. 100 bis 700 bp zerlegt, die von der Größe her für eine Subklonierung und Sequenzierung geeignet sind. Diese Eco RI/Pst I-Fragmente wurden in präparativem Maßstab isoliert (ausgenommen die zweite Bande in Spur 3 im Bereich von ca. 800 bp, die einem Fragment des pBR322 entspricht) und in die M13-Vektoren tg130 und tg131 eingesetzt. Eine erste Sequenzanalyse des Subklons ma2 (Abb.11) vom 5'-Ende aus ergab, daß es sich hier tatsächlich um $\alpha 2$ (IV)-Sequenzen handelte, und zwar aus dem tripelhelikalen Bereich. Dies wurde durch Vergleich der von der cDNA abgeleiteten Peptidsequenzen mit den durch Proteinsequenzierung bekannten Partialsequenzen der $\alpha 2$ (IV)-Tripelhelix (siehe Abb.18, S.62) festgestellt. Nachdem also geklärt war, daß pAIIa $\alpha 2$ (IV)-Sequenzen enthält, wurde dieser Klon vollständig sequenziert. Die drei Subklone ma2, ma3 und ma4 waren in beiden Orientierungen in die Vektoren M13tg130 und tg131 eingesetzt und wurden so weit wie möglich sequenziert (Abb.11). Das dem Subklon ma1 entsprechende Fragment wurde nur in den M13tg131 integriert (zur Sequenzierung des komplementären Stranges), da man damit rechnete, daß eine Sequenzierung des codierenden Stranges wegen der GC-Schwänze am 5'-Ende dieses Fragments nicht durchführbar sein würde. Es zeigte sich allerdings bei der Sequenzanalyse von ma1, daß hier die GC-Schwänze (wie auch bei dem später isolierten Klon pAIIb, Abb.11) sehr kurz (13 bis 14 bp) sind und die Spezifität der DNA-Polymerase bei der Sequenzierung nicht beeinträchtigen.

Die versuchte Subklonierung des ca. 100 bp langen Pst I-Fragments am 3'-Ende von pAIIa schlug leider fehl.

Da die Subklone ma2 und ma4 sehr lang sind (630 bp und 700 bp), war eine durchgehende Sequenzierung beider Stränge nicht möglich. Daher mußten weitere Subklone konstruiert werden (ma5 bis ma7). Hierzu wurde das ma4 entsprechende Eco RI/Pst I-Fragment mit Alu I inkubiert und das 630 bp-Fragment (ma2) mit

Bam HI. Die entstandenen Teilfragmente wurden ebenfalls in M13 eingesetzt, ausgenommen das Eco RI/Bam HI- Fragment von ma2. Hier konnte auf eine Subklonierung und Sequenzierung verzichtet werden, da inzwischen ein in diesem Bereich überlappender Klon (pAIIb, Abb.11) isoliert worden war, auf den später eingegangen wird.

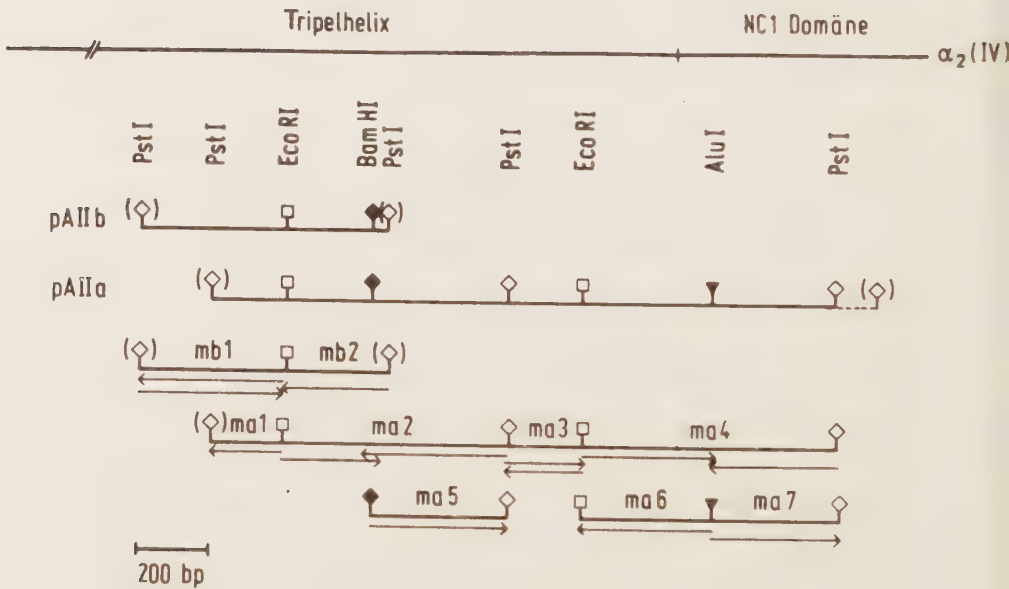


Abb.11: Restriktionskarte der Klone pAIIa und pAIIb und Sequenzierungsstrategie
Richtung und Länge der durchgeführten Sequenzierungen sind durch Pfeile gekennzeichnet. ma1 bis ma7 sind Subklone von pAIIa in den M13-Vektoren tg130 und tg131, mb1 und mb2 entsprechend von pAIIb. Die eingeklammerten Pst I-Stellen werden vom Vektor pBR322 beige-tragen.

Die Sequenzierung von Klon pAIIa ergab, daß sich die von der cDNA abgeleiteten Proteinsequenzen aus 430 Aminosäuren C-terminaler tripelhelikaler Sequenz und 145 Aminosäuren der globulären Domäne zusammensetzen. Die NC1-Sequenz ist nicht vollständig in Klon pAIIa enthalten.

Wie bereits in der Einleitung erwähnt, waren auf Proteinebene Sequenzen tripelhelikaler Peptidfragmente der $\alpha 2$ -Kette vorhanden, deren Reihenfolge jedoch nicht klar war. Durch Sequenzierung des Klons pAIIa konnten die 430 C-terminalen Aminosäuren der Tripelhelix bestimmt werden.

Um hier noch weitergehende Informationen über die Primärstruktur der Tripelhelix zu erhalten, sollte zunächst versucht werden, mit Hilfe von Klon pAIIa 5'-überlappende Klone zu identifizieren. Zu diesem Zweck wurde das Eco RI/Pst I-Fragment, das ma2 in Abb.11 entspricht, radioaktiv markiert und wiederum gegen die EHS-Genbank hybridisiert. Dies führte zur Isolierung von Klon pAIIb. Eco RI/Pst I-Fragmente von pAIIb wurden ebenfalls in die M13-Vektoren tg130 und tg131 subkloniert (mb1, in beiden Orientierungen, und mb2) und sequenziert (Abb.11). Die Subklone mb1 und mb2 konnten ausgehend von den Enden der cDNA sequenziert werden, da, wie bereits erwähnt, die kurzen GC-Schwänze die Spezifität der Reaktionen nicht herabsetzten. pAIIb überlappt ca. 500 bp mit pAIIa und trägt lediglich ca. 200 bp bisher unbekannter tripelhelikaler Sequenz bei.

Das Autoradiogramm eines Sequenzgels (Abb.12) zeigt das typische Bandenmuster tripelhelikaler Sequenzen. Die in konstanten Abständen wiederkehrenden Codons für Glycin, GGN, führen zur Ausprägung einer kontinuierlichen Leiter in der G-Spur bei Sequenzierung des codierenden Stranges, bzw. in der C-Spur bei Sequenzierung des komplementären Stranges. Die Besonderheit des Typ IV Kollagens, Unterbrechungen der tripelhelikalen Sequenz, sind hier als Diskontinuitäten im charakteristischen Bandenmuster leicht zu erkennen.

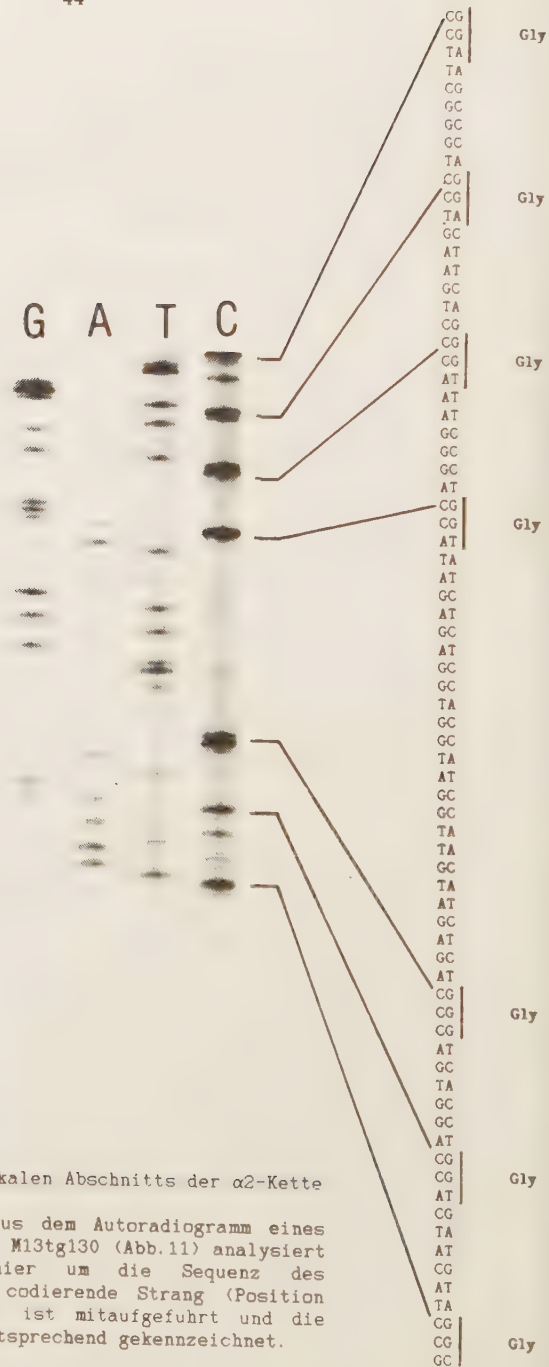


Abb. 12: Sequenzierung eines tripehelikalen Abschnitts der $\alpha 2$ -Kette mit einer Unterbrechung
Gezeigt ist ein Ausschnitt aus dem Autoradiogramm eines Sequenzgels, wo Subklon *ma2* im M13tg130 (Abb. 11) analysiert wurde. Es handelt sich hier um die Sequenz des komplementären Stranges. Der codierende Strang (Position 856 bis 930 der Tripehelix) ist mitaufgeführt und die Codons für Glycin sind hier entsprechend gekennzeichnet.

Gemeinsam codieren pAIIa und pAIIb für 511 Aminosäuren der C-terminalen tripelhelikalen Region von $\alpha 2(IV)$. Es erschien wenig aussichtsreich nach Klonen zu suchen, die noch weiter in die Tripelhelix reichen, da das 5'-Ende von Klon pAIIb schon über 3 kb vom Start der cDNA-Synthese entfernt liegt. Die Möglichkeit, daß die Reverse Transkriptase bei der Synthese des ersten Stranges der cDNA noch weiter vordringen konnte, wird mit zunehmender Entfernung vom 3'-Ende der RNA unwahrscheinlicher, zumal vermutlich die Ausbildung von Sekundärstrukturen im GC-reichen tripelhelikalen Abschnitt der mRNA häufig zu Abbrüchen führt.

Somit konzentrierte sich die weitere Arbeit darauf, die Sequenz der NC1-Domäne der $\alpha 2$ -Kette zu vervollständigen.

4.2. Isolierung und Charakterisierung von Klon pAIIc

Nachdem keiner der bisher aus der EHS-cDNA-Genbank isolierten Klone (p5, p9, pAIIa und pAIIb) das 3'-Ende der NC1-Sequenz, bzw die nichttranslatierte Region enthielt - die vermuteten Ursachen hierfür wurden bereits genannt (Teil I,3.2., S. 37) - wurden für die weitere Isolierung von Typ IV-spezifischen Klonen auf andere Weise konstruierte cDNA-Genbanken eingesetzt. Diese Genbanken (cI, cII und PG3) werden in Teil III,1. eingehend beschrieben. Vorweggenommen sei, daß diese Genbanken analog zur Okayama-Berg-Methode (1982) synthetisiert wurden, und daher alle integrierten cDNAs den 3'-Bereich enthalten sollten. Zur Vervollständigung der NC1-Sequenzen von $\alpha 2(IV)$ wurde die Genbank cII eingesetzt (siehe Teil III,1., Abb.21, S. 86).

Ein Klon mit dem fehlenden Ende der NC1-Sequenz von $\alpha 2(IV)$ wurde durch "screening" der Genbank mit einem 3'-gelegenen Fragment von Klon pAIIa (ma4 in

Abb.11) isoliert. Dieses Eco RI/Pst I-Fragment wurde durch Nicktranslation radioaktiv markiert und als Hybridisierungsprobe gegen die Genbank cII eingesetzt. Von den wenigen positiven Klonen, die diese Hybridisierung brachte, wurde der Längste (ca. 1700 bp), genannt pAIIc, weiter charakterisiert. Bei einer Restriktionskartierung zeigte sich (Abb.13), daß pAIIc eine Pst I-Stelle im 5'-Bereich enthält. Hier konnte man davon ausgehen, daß diese Position der 3'-gelegenen Pst I-Stelle von Klon pAIIa (Abb.13) entspricht. Da die Sequenz des Überlappungsbereichs (ca. 440 bp) der beiden Klone bereits bekannt war, wurden nur die 3' dieser Pst I-Stelle folgenden Sequenzen von pAIIc subkloniert. Die zwei Pst I/Ava I-Fragmente mc1 und mc2 wurden in beiden Orientierungen in M13tg130 und M13tg131 integriert und vollständig sequenziert. mc1 enthält die fehlenden 150 bp der NC1-Sequenz, deren Ende durch ein Stopcodon angezeigt wird, und den Beginn der nichttranslatierten Region der $\alpha 2$ -Kette. Bei einem dritten Pst I/Ava I-Fragment (mc3) konnte nur der codierende Strang analysiert werden, da die Sequenzierung des komplementären Stranges vom 3'-Ende her aufgrund des dort lokalisierten ca. 150 bp langen Poly(A)-Schwanzes nicht durchführbar war. Um die Sequenzdaten aus der Analyse von mc3 so weit als möglich abzusichern, bzw. zu vervollständigen, wurden weitere Subklone (mc4 bis mc6) konstruiert und sequenziert.

Klon pAIIc vervollständigt die NC1-Sequenz der $\alpha 2$ -Kette, die damit insgesamt 681 bp (227 Aminosäuren) umfaßt. Außerdem enthält Klon pAIIc ca. 950 bp der nichttranslatierten Region, sowie ca. 150 bp des Poly(A)-Schwanzes am 3'-Ende. Trotz dieser Präsenz des Poly(A)-Schwanzes ist nicht die gesamte nichttranslatierte Region von $\alpha 2(IV)$ vertreten - ein Phänomen, das in Teil III (S. 89) behandelt werden wird.

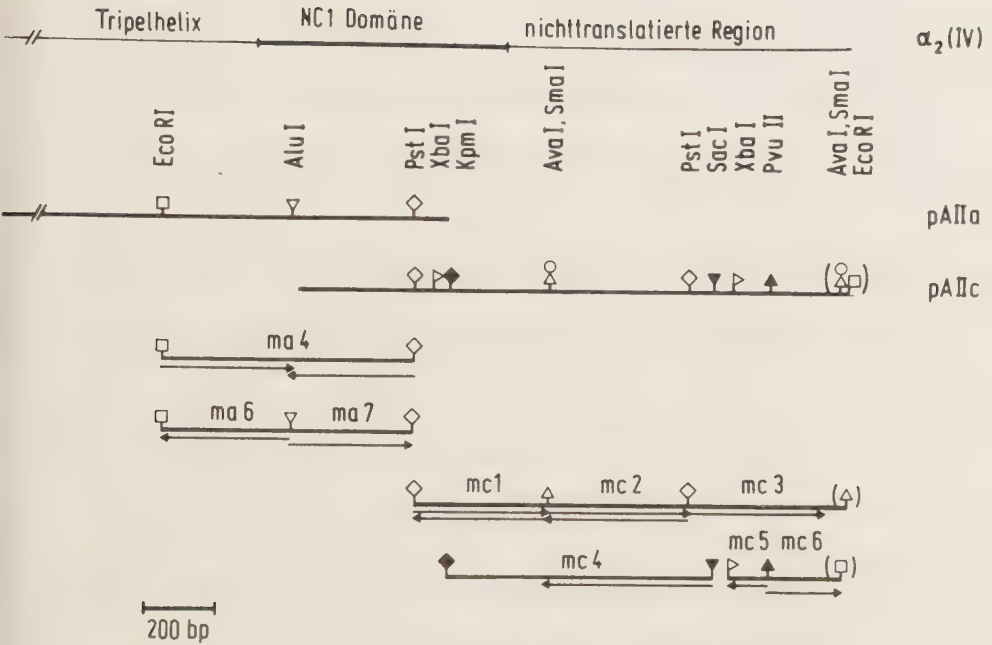


Abb. 13: Restriktionskarte von Klon pAIIc und Sequenzierungsstrategie. Der C-terminale Bereich des überlappenden Klon pAIIa mit seinen Subklonen ma4, ma6 und ma7 ist mitaufgeführt. mc1 bis mc6 benennt die M13-Subklone (tg130 und tg131) von Klon pAIIc. Die sequenzierten Bereiche sind durch Pfeile angezeigt, die eingeklammerten Restriktionsstellen kennzeichnen Restriktionsstellen aus der Polylinkerregion des Vektors.

Teil II: Charakterisierung der NC1-Domäne und der Tripelhelix auf Protein- und auf DNA-Ebene

1. Die Struktur der NC1-Domäne

1.1. DNA-Sequenzen und abgeleitete Proteinsequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$

Der Ausgangspunkt für die Sequenzanalysen der cDNA der NC1-Region waren kurze Teilsequenzen des Proteins, deren Lokalisation innerhalb des NC1-Bereichs allerdings nicht bekannt war. So waren zu Beginn dieser Arbeit durch Proteinsequenzierung für Mensch 44 Aminosäuren der $\alpha 1$ -Kette und 54 Aminosäuren der $\alpha 2$ -Kette aufgeklärt. Für $\alpha 1(IV)$ stand zusätzlich die Sequenz eines 50 Aminosäuren langen Abschnitts von Maus zur Verfügung, die von einem entsprechenden cDNA-Bereich abgeleitet werden konnte (Kurkinen, nicht veröffentlicht). Parallel zu den molekularbiologischen Arbeiten über den NC1-Bereich von Maus wurde die proteinchemische Sequenzierung der NC1-Region von $\alpha 1(IV)$ und $\alpha 2(IV)$ von Mensch weitergeführt (Siebold, Dissertation 1986). Durch den Vergleich der Sequenzdaten war eine gute Kontrollmöglichkeit gegeben. Außerdem war der Homologievergleich der NC1-Sequenzen von Mensch und Maus von Interesse.

Wie in Teil I beschrieben, wurden cDNA-Klone isoliert, die die vollständigen Sequenzen der NC1-Region von $\alpha 1(IV)$ und $\alpha 2(IV)$ der Maus abdecken. In Abbildung 14 sind die entsprechenden DNA-Sequenzen und die davon abgeleiteten Proteinsequenzen für beide Ketten dargestellt.

Die NC1-Region der $\alpha 1$ -Kette (Abb.14a) setzt sich aus 687 Basen (229 Aminosäuren) zusammen. Das Ende der codierenden Sequenz wird hier durch zwei Stop-codons in kurzer Folge (siehe nichttranslatierte Region, Abb.25, S. 96) angezeigt. Dagegen wird das 3'-Ende der $\alpha 2$ -Sequenz (Abb.14b) nur durch ein Stop-codon markiert. Die NC1-Region von $\alpha 2(IV)$ umfaßt 681 Basen (227 Aminosäuren) und ist damit um zwei Aminosäuren kürzer als die NC1-Region der $\alpha 1$ -Kette.

$\alpha 1$ (IV)

1	TCTGTGGACCATGGCTTCCTTGTGACCAGGCACAGTCAGACAACAGATGACCCACTGTGT	60
1	S V D H G F L V T R H S Q T T D D P L C	20
	I Q	
61	CCCCAGGGACCAAAATTCTTTACCATGGATACTCTCTGCTCTATGTTCAAGGCAACGAG	120
21	P P G T K I L Y H G Y S L L Y V Q G N E	40
121	CGTGCCACGGGCAGGACTTGGGTACGGCTGGCAGCTGCCTGCGTAAGTTCAGCACCATG	180
41	R A H G Q D L G T A G S C L R K F S T M	60
181	CCCTTCTCTCTCTGCAACATCAACAACGCTCTGCAACTTCGCCTCCAGGAACGACTACTCT	240
61	P F L F C N I N N V C N F A S R N D Y S	80
241	TACTGGCTGTCCACGCCAGAGCCCATGCCCATGTCCATGGCACCCATCTCTGGGGACAAC	300
81	Y W L S T P E P M P M S M A P I S G D N	100
	T E	
301	ATCCGGCCCTTCATTAGCAGGTGTGCGGTTTGTGAAGCACCGGCCATGGTGATGGCGGTA	360
101	I R P F I S R R C A V C E A P A M V M A V	120
361	CACAGTCAGACCATTAGATTCCGAGTGCCTTAACGGTTGGTCTCTACTGTGGATCGGC	420
121	H S Q T I Q I P Q C P N G W S S L W I G	140
421	TATTCCTTCGTGATGCACACCAGCGCTGGTGCTGAAGGTTCCGGCCAAGCCCTCGCATCC	480
141	Y S F V M H T S A G A E G S G Q A L A S	160
481	CCCGGCTCCTGTCTGGAAGAGTTAGAAGCGCCCCATTATCGAGTGCCACGGCAGAGGA	540
161	P G S C L E E F R S A P F I E C H G R G	180
541	ACGTGCAATTACTACGCAAATGCTTACAGCTTTTGGCTCGCCACCATAGAGAGAAGCGAG	600
181	T C N Y Y A N A Y S F W L A T I E R S E	200
601	ATGTTCAAGAAGCCCCACGCCGTCCACCTTGAAGGCAGGGGAGCTGCGCACACACGTGAG	660
201	M F K K P T P S T L K A G E L R T H V S	220
661	CGCTGCCAAGTGTGCATGAGAAGAACATAA	690
221	R C Q V C M R R T *	229

Abb. 14a: cDNA-Sequenzen und abgeleitete Proteinsequenzen des NC1-Bereichs von $\alpha 1$ (IV). Identische Proteinsequenzen bei $\alpha 1$ (IV) von Mensch (Siebold, Dissertation, 1986) sind durch eine Linie gekennzeichnet. Abweichungen dieser Humansequenzen gegenüber den Mausequenzen sind jeweils in der dritten Zeile aufgeführt.

$\alpha 2(IV)$

1	GTGAGCATCGGCTACCTCCTGGTGAAGCACAGCCAAACGGACCAGGAACCCATGTGCGCCT	60
1	V S I G Y L L V K H S Q T D Q E P M C P	20
61	GTGGGCATGAACAAGCTCTGGAGTGGGTACAGCCTGCTATATTTTGAGGGCCAGGAGAAA	120
21	V G M N K L W S G Y S L L Y F E G Q E K	40
121	GCGCACAACCAGGACCTAGGACTGGCAGGCTCCTGCCTGGCACGCTTCAGCACCATGCCT	180
41	A H N Q D L G L A G S C L A R F S T M P	60
181	TTCCTGTACTGCAATCCGGGTGACGCTCTGCTACTATGCCAGCCGCAACGACAAGTCCTAC	240
61	F L Y C N P G D V C Y Y A S R N D K S Y	80
241	TGGCTCTCCACCACGGCCCCTCTGCGCATGATGCCCTGGCTGAGGAGGAAATCAAGCCC	300
81	W L S T T A P L P M M P L A E E E I K P	100
301	TACATCAGCCGCTGCTCTGTGTGCGAGGCTCCGGCCGTGGCCATTGCCGTGCACAGCCAG	360
101	Y I S R C S V C E A P A V A I A V H S Q	120
361	GATACCTCTATACCCCACTGCCCGGTGGGTGGCGGAGTTTGTGGATCGGATATTCATTC	420
121	D T S I P H C P A G W R S L W I G Y S F	140
421	CTCATGCACACTGCAGCCGGGATGAAGGCGGTGGCCAGTCACTGGTGTGCGCCGGGCGAGC	480
141	L M H T A A G D E G G G Q S L V S P G S	160
481	TGTCTAGAGGACTTCCGTGCAACGCCATTATCGAGTGTAACGGGGGCCGTGGTACCTGC	540
161	C L E D F R A T P F I E C N G G R G T C	180
541	CACTACTTCGCTAACAAAGTACAGCTTCTGGCTGACCACGATCCCAGAGCAGAACTTCCAG	600
181	H Y F A N K Y S F W L T T I P E Q N F Q	200
601	AGCACACCATCCGCTGACACGCTCAAGGCTGGCCTCATCCGCACGCACATCAGCCGCTGC	660
201	S T P S A D T L K A G L I R T H I S R C	220
661	CAAGTGTGCATGAAGAATCTGTGA	684
221	Q V C M K N L *	227

Abb. 14b: cDNA-Sequenzen und abgeleitete Proteinsequenzen des NC1-Bereichs von $\alpha 2(IV)$. Identische Proteinsequenzen bei $\alpha 2(IV)$ von Mensch (Siebold, Dissertation, 1986) sind durch eine Linie gekennzeichnet. Abweichungen dieser Humansequenzen gegenüber den Mausequenzen sind jeweils in der dritten Zeile aufgeführt.

In Abbildung 14a und 14b sind die Humanproteinsequenzen für $\alpha 1(IV)$ und $\alpha 2(IV)$ mitaufgeführt, und zwar markiert eine Linie die Abschnitte, die bei Mensch und Maus identisch sind. Hier zeigen sich hochgradige Homologien mit nur ganz vereinzelt auftretenden Austauschen.

Bei der $\alpha 1$ -Kette konnten durch Proteinsequenzierung 178 Aminosäuren der NC1-Domäne von Mensch bestimmt werden (Siebold, Dissertation 1986). Nur an vier Positionen (Pos. 15, 19, 97 und 99) sind hier Unterschiede zur Maussequenz zu finden, wobei es sich zudem in zwei Fällen (Pos. 97 und 99) um Austausche verwandter Aminosäuren handelt.

Durch die Arbeit von Pihlajaniemi et al. (1985) ist mittlerweile die gesamte Proteinsequenz der humanen NC1-Region von $\alpha 1(IV)$ bekannt (abgeleitet von cDNA). Sie ist, wie der entsprechende Bereich der Maus, 229 Aminosäuren lang und weist neben den bereits angeführten vier Austauschen an drei weiteren Aminosäurepositionen Differenzen zur Maussequenz auf. Somit sind 97 % der Aminosäuren in der NC1-Region der $\alpha 1$ -Kette von Mensch und Maus identisch.

Für $\alpha 2(IV)$ konnten durch Sequenzierung von Peptidfragmenten 208 Aminosäuren des NC1-Bereichs der humanen $\alpha 2$ -Kette bestimmt werden (Siebold, Dissertation 1986). Die entsprechenden Sequenzabschnitte sind in Abbildung 14b markiert. Hier treten an elf Positionen Unterschiede zwischen beiden Sequenzen auf, wobei es sich in sechs Fällen um Austausche verwandter Aminosäuren handelt. Somit sind 94,7 % der Aminosäuren des NC1-Bereichs von $\alpha 2(IV)$ bei Mensch und Maus identisch - ein Prozentsatz, der etwas niedriger liegt als der entsprechende Wert (97 %) von $\alpha 1(IV)$.

In der Einleitung wurde bereits erwähnt, daß Typ IV Kollagen-Moleküle nicht wie die interstitiellen Prokollagene proteolytisch prozessiert werden. Man ging davon aus, daß, wenn überhaupt, nur sehr kurze Propeptide vor der Quervernetzung der Moleküle abgespalten werden. Durch die vorliegenden Sequenzdaten war es nun möglich, diese Frage zumindest in Bezug auf den C-Terminus eindeutig zu klären. So stimmten die durch Proteinsequenzierung erhaltenen C-terminalen Enden von $\alpha 1(IV)$ und $\alpha 2(IV)$ aus quervernetztem Typ IV Kollagen (Siebold, Dissertation 1986) mit den von der cDNA abgeleiteten Endbereichen überein. Dadurch wurde gezeigt, daß das carboxyterminale Ende der Moleküle beim Einbau in das Netzwerk intakt bleibt.

Aus der Aminosäurezusammensetzung von $\alpha 1(IV)$ und $\alpha 2(IV)$ innerhalb der NC1-Domäne lassen sich Molekulargewichte von ca. 25300 ($\alpha 1$) und 25000 ($\alpha 2$) errechnen. In dieser Größenordnung liegen auch die auf Proteinebene ermittelten relativen Molekulargewichte von isolierten NC1-Monomeren. Hier ergaben Gleichgewichtszentrifugation, Elektronenmikroskopie und SDS-Gelelektrophorese einen Durchschnittswert von 28000 ± 3000 Da, einschließlich eines Carbohydratanteils von 3 bis 6 % (Weber et al., 1984).

Die Carbohydratanalyse hatte die Gegenwart von Glucosamin und Galactosamin angezeigt (Weber et al., 1984). Daher ging man davon aus, daß es in der NC1-Domäne sowohl O-glykosidisch an Serin und Threonin, als auch N-glykosidisch an Asparagin gebundene Carbohydratseitenketten gibt. Die Sequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$ enthalten jedoch keine Bindungsstellen (Asp-X-Ser und Asp-X-Thr) für N-glykosidisch gebundene Zucker. Der Nachweis von Glucosamin als Bestandteil der NC1-Domäne dürfte daher die Folge einer Kontamination gewesen sein. Auch für O-glykosidisch gebundene Zucker gibt es keinen direkten Hinweis. Bei der Aminosäuresequenzierung zeigte sich bei keiner der mit Serin und Threonin besetzten Positionen eine Lücke, wie es bei einer vollständigen Glykosylierung zu erwarten wäre.

1.2. Homologie von $\alpha 1(IV)$ und $\alpha 2(IV)$ im Bereich der NC1-Domäne

Die Daten aus den proteinchemischen Analysen der NC1-Domäne ließen schon früh erkennen, daß in diesem Bereich eine starke Homologie von $\alpha 1$ - und $\alpha 2$ -Kette ausgeprägt ist. Der direkte Vergleich der beiden vollständigen Sequenzen verdeutlicht dieses Charakteristikum der globulären Domäne (Abb.15).

Um eine optimale Zuordnung der homologen Abschnitte zu erreichen, mußten drei Lücken in die $\alpha 2$ -Sequenz (Positionen 1, 92 und 94) und eine Lücke in die $\alpha 1$ -Sequenz (Position 179) eingefügt werden. Nach dieser Anpassung zeigt sich, daß 144 von 230 korrespondierenden Positionen in beiden Ketten von identischen Aminosäuren besetzt werden. Das entspricht einem Anteil von 62.6 %. Die Ähnlichkeit beider Ketten wird noch erhöht durch das Auftreten von verwandten Aminosäuren (nach "Simplify", Computerprogramm UWGCG; Devereux et al., 1984) an gleichen Positionen (17.7 %). Auffallend ist die Tatsache, daß in beiden α -Ketten innerhalb der NC1-Domäne zwölf Cysteine vorkommen, und daß alle diese Cysteinreste in $\alpha 1(IV)$ und $\alpha 2(IV)$ an den gleichen Stellen auftreten (in Abb.15 durch Sternchen markiert). Die Konservierung dieser Cysteine ist hinsichtlich der Ausbildung von Disulfidbrücken innerhalb der NC1-Domäne von Bedeutung, die zur Stabilisierung von Sekundär- und Tertiärstrukturen beiträgt (siehe Diskussion). Auch weitere, für die Konformation von Peptidketten entscheidende Strukturmerkmale wie Hydropathie (Verteilung hydrophiler und hydrophober Aminosäuren), sowie die Lokalisation von Bereichen, die für die Ausbildung von α -Helix, β -Faltblatt oder Knicken im Molekül befähigt sind, stimmen bei den NC1-Sequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$ weitgehend überein (Siebold, Dissertation 1986).

$\alpha 1$ (IV)	1	S	V	D	H	G	F	L	V	T	R	H	S	Q	T	T	D	D	P	L	C	*	20
$\alpha 2$ (IV)	1	(S)	V	S	I	G	Y	L	L	V	K	H	S	Q	T	D	Q	E	P	M	C	*	19
	21	P	P	G	T	K	I	L	Y	H	G	Y	S	L	L	Y	V	Q	G	N	E		40
	20	P	V	G	M	N	K	L	W	S	G	Y	S	L	L	Y	F	E	G	Q	E		39
	41	R	A	H	G	Q	D	L	G	T	A	G	S	C	L	R	K	F	S	T	M		60
	40	K	A	H	N	Q	D	L	G	L	A	G	S	C	L	A	R	F	S	T	M		59
	61	P	F	L	F	C	N	I	N	N	V	C	N	F	A	S	R	N	D	Y	S		80
	60	P	F	L	Y	C	N	P	G	D	V	C	Y	Y	A	S	R	N	D	K	S		79
	81	Y	W	L	S	T	P	E	P	M	P	M	S	M	A	P	I	S	G	D	N		100
	80	Y	W	L	S	T	T	A	P	L	P	M	-	M	-	P	L	A	E	E	E		97
	101	I	R	P	F	I	S	R	C	A	V	C	E	A	P	A	M	V	M	A	V		120
	98	I	K	P	Y	I	S	R	C	S	V	C	E	A	P	A	V	A	I	A	V		117
	121	H	S	Q	T	I	Q	I	P	Q	C	P	N	G	W	S	S	L	W	I	G		140
	118	H	S	Q	D	T	S	I	P	H	C	P	A	G	W	R	S	L	W	I	G		137
	141	Y	S	F	V	M	H	T	S	A	G	A	E	G	S	G	Q	A	L	A	S		160
	138	Y	S	F	L	M	H	T	A	A	G	D	E	G	G	G	Q	S	L	V	S		157
	161	P	G	S	C	L	E	E	F	R	S	A	P	F	I	E	C	H	G	-	R		179
	158	P	G	S	C	L	E	D	F	R	A	T	P	F	I	E	C	N	G	G	R		177
	180	G	T	C	N	Y	Y	A	N	A	Y	S	F	W	L	A	T	I	E	R	S		199
	178	G	T	C	H	Y	F	A	N	K	Y	S	F	W	L	T	T	I	P	E	Q		197
	200	E	M	F	K	K	P	T	P	S	T	L	K	A	G	E	L	R	T	H	V		219
	198	N	F	Q	S	T	P	S	A	D	T	L	K	A	G	L	I	R	T	H	I		217
	220	S	R	C	Q	V	C	M	R	R	T												229
	218	S	R	C	Q	V	C	M	K	N	L												227

Abb.15: Zuordnung der Proteinsequenzen des NC1-Bereichs von $\alpha 1$ (IV) und $\alpha 2$ (IV) von Maus

Identische Aminosäurepositionen in beiden Ketten sind eingerahmt, homologe Aminosäuren durch Punkte gekennzeichnet. Die Lokalisationen der Cysteine sind durch Sternchen markiert.

Der Serinrest in Pos.1 von $\alpha 2$ (IV) ist eingeklammert, da er - im Gegensatz zu dem entsprechenden Serin von $\alpha 1$ (IV) - als Bestandteil des letzten Triplets (G R S) der tripelhelikalen Sequenz von $\alpha 2$ (IV) gelten kann. Dieses Serin wird daher auch der tripelhelikalen Region zugerechnet (Abb.18), obwohl es ebenso als erste Aminosäure des NC1-Bereichs von $\alpha 2$ (IV) - homolog zu dem ersten Serin von $\alpha 1$ (IV) - angesehen werden konnte.

Die Homologie von $\alpha 1(IV)$ und $\alpha 2(IV)$ innerhalb der globulären Domäne spiegelt sich auch in der Aminosäurezusammensetzung wieder (Tab.5). Die nach den Eigenschaften der Aminosäureseitengruppen geordnete Auflistung (nach "Simplify", UWGCG; Devereux et al., 1984) verdeutlicht außerdem den hydrophoben Charakter der NC1-Domäne (Anteil hydrophober Aminosäuren: 30 % bei $\alpha 1(IV)$, 31 % bei $\alpha 2(IV)$).

Tab.5: Aminosäurezusammensetzung von $\alpha 1(IV)$ und $\alpha 2(IV)$ in der NC1-Domäne

Aminosäuren	$\alpha 1(IV)$		$\alpha 2(IV)$	
	Anzahl	Prozentsatz	Anzahl	Prozentsatz
P, A, G, S, T	89	38.8 %	85	37.4 %
Q, N, E, D	34	14.8 %	36	15.8 %
H, K, R	26	13.3 %	23	10 %
L, I, V, M	44	19.2 %	46	20.2 %
F, Y, W	24	10.5 %	25	11 %
C	12	5.2 %	12	5.3 %

1.3. Interne Homologie der NC1-Domäne und Homologie zu den Carboxypropeptiden der interstitiellen Kollagene

Das überraschendste Ergebnis der Analyse der Primärstruktur der NC1-Domäne von Typ IV Kollagen war, daß innerhalb beider α -Peptidketten eine interne Homologie ausgebildet ist. Eine optimale Anordnung der homologen Bereiche wurde durch Anwendung des Computerprogrammes "Align" (Barker et al., 1983) erreicht (Abb.16). Jeweils die erste Hälfte der NC1-Sequenz (NC1A) ist homolog zur zweiten Hälfte (NC1B). Bei der $\alpha 1$ -Kette betrifft dies die Positionen 2 bis 111 (NC1A) und

112 bis 225 (NC1B), bei der $\alpha 2$ -Kette die Positionen 1 bis 108 (NC1A) und 109 bis 223 (NC1B). Bemerkenswert ist, daß auch hier alle Cysteinreste (je sechs pro Hälfte) korrespondierende Positionen besetzten (Abb.16). Einschließlich dieser Cysteine sind innerhalb beider Homologiehälften bei beiden Ketten jeweils 35 % der Aminosäuren identisch, weitere 20 % ähnlich. Bei der Zuordnung der homologen Hälften von $\alpha 2$ (IV) erscheinen jedoch mehr Lücken (fünfzehn), die als Deletionen interpretiert werden müssen, als bei $\alpha 1$ (IV) (nur elf).

Es ist auffällig, daß die identischen Aminosäuren in beiden Hälften beider Ketten in Blöcken auftreten, zwischen denen weniger konservierte Bereiche liegen. Diese Blockbildung trifft insbesondere auf die flankierenden Sequenzen der Cysteine zu. Auch zwei Abschnitte mit benachbarten aromatischen Aminosäuren (vor allem Tyrosin und Tryptophan) sind als Block in allen vier Hälften konserviert (jeweils nach dem ersten und dem vierten Cystein).

Die NC1-Domäne sitzt wie die Carboxypropeptide der interstitiellen Kollagene am C-terminalen Ende des Moleküls und scheint wie sie eine wichtige Rolle bei der richtigen Zusammenlagerung dreier α -Ketten vor der Tripelhelixbildung zu spielen (siehe Diskussion). Es stellte sich daher die Frage, inwieweit zwischen diesen globulären C-terminalen Domänen von Typ IV Kollagen und den Kollagenen Typ I bis Typ III eine Ähnlichkeit besteht.

Stellvertretend für die interstitiellen Kollagene, die untereinander sehr homolog sind, wurde das Carboxypropeptid der $\alpha 1$ -Kette von Typ III Kollagen (Huhn) auf Sequenzhomologien zu der NC1-Region von Typ IV Kollagen hin analysiert. Es zeigte sich, daß, bis auf einen kurzen Abschnitt (Position 75 bis 107 des Carboxypropeptids, nach Yamada et al., 1983) keinerlei Sequenzhomologien bestehen. Dieses 33 Aminosäuren lange Segment allerdings konnte aufgrund identischer und verwandter Aminosäuren homologen Abschnitten beider NC1-Hälften

$\alpha 1(IV)$

NC1(A)	1	S V D H G F L V T R H S Q T T D D P L C P P G T K
NC1(B)	112	E A P A M V M A V H S Q T I Q I P Q C P N G W S
NC1(A)	26	I L Y H G Y S L L - Y V Q G N E R A H G Q D L G T
NC1(B)	136	S L W I G Y S F V M H T S A G A E G S G Q A L A S
NC1(A)	50	A G S C L R K F S T M P F L F C N I N N V C N F A
NC1(B)	161	P G S C L E E F R S A P F I E C H G R G T C N Y Y
PRO III		R K N P A R N C R D L K F C H P E
NC1(A)	75	S R N D Y S Y W L S T P E P M P M S M A P - - -
NC1(B)	186	A N A Y S F W L A T I E R S E M F K K P T P S T
PRO III		L K S G - E Y W I D P N Q G C K M
NC1(A)	96	I S G D N I R P F I S R C A V C - - -
NC1(B)	210	L K A G E L R T H V S R C Q V C M R R T

$\alpha 2(IV)$

NC1(A)	1	V S I G Y L L V K H S Q T D Q E P M C P V G M N K
NC1(B)	109	E A P A V A I A V H S Q D T S I P H C P A G W R S
NC1(A)	26	L W S G Y S - L L Y F E G Q E K A H N Q D L G L A
NC1(B)	134	L W I G Y S F L M H T A A G D E G G Q S L V S P
NC1(A)	50	G S C L A R F S T M P F L Y C N P G - D V C - Y Y
NC1(B)	159	G S C L E D F R A T P F I E C N G G R G T C H Y F
PRO III		R K N P A R N C R D L K - F C H P E
NC1(A)	73	A S R N D K S Y W L S T T A P L P M M P L A E E E
NC1(B)	184	A N K Y - - S F W L T T I P E Q N F Q S T P S A D
PRO III		L - K S G - E Y W I D P N Q G C K M
NC1(A)	98	- - - - I K P Y I S R C S V C - - -
NC1(B)	207	T L K A G L I R T H I S R C Q V C M K N L

Abb. 16: Interne Homologie von $\alpha 1(IV)$ und $\alpha 2(IV)$ im NC1-Bereich und Homologie zum Carboxypropeptid von $\alpha 1(III)$ (Yamada et al., 1983)
Identische Aminosäuren sind durch eine Linie eingerahmt, homologe Aminosäuren durch Punkte

in beiden α -Ketten zugeordnet werden (Abb.16). Sogar zwei Cysteinreste (Position 82 und 88) des $\alpha 1$ (III) Carboxypropeptids besetzen Positionen, die auch in den beiden homologen Hälften der NC1-Sequenzen von α (IV) und $\alpha 2$ (IV) von Cysteinen eingenommen werden. Ferner ist einer der Bereiche mit benachbarten aromatischen Aminosäuren konserviert (Pos. 97 und 98 von $\alpha 1$ (III)). Diese Befunde sind insofern bemerkenswert, als die betreffende 33 Aminosäuren lange Sequenz zu der mit am meisten konservierten Region der Carboxypropeptide der interstitiellen Kollagene gehört und offensichtlich auch bis zu einem gewissen Grad im C-terminalen Bereich des Typ IV Kollagens präsent ist, und das sogar zweimal pro Kette. Auf die mögliche Bedeutung dieser konservierten Bereiche hinsichtlich der Funktionen der NC1-Domäne und der Carboxypropeptide der Kollagene Typ I bis III wird im Diskussionsteil eingegangen.

Strukturähnlichkeiten zwischen der NC1-Domäne und den Carboxypropeptiden der interstitiellen Kollagene sind also nur sehr begrenzt entwickelt. Eine Gemeinsamkeit stellt jedoch die hochgradige Konservierung dieser C-terminalen Regionen bei den verschiedenen α -Ketten dar. Bei den beiden α -Ketten von Typ I Kollagen von Huhn (Fuller und Boedtker, 1981; Dickson et al., 1981) sind im Bereich der Carboxypropeptide (bei $\alpha 1$ (I) 247, bei $\alpha 2$ (I) 246 Aminosäuren) 61 % der korrespondierenden Positionen identisch, weitere 22.2 % homolog. Der Vergleich zu den entsprechenden Werten der NC1-Domäne (62.6 % identisch, 17.7 % homolog) zeigt, daß diese Regionen in gleichem Ausmaß konserviert sind.

2. Die Struktur der Tripelhelix

2.1. DNA-Sequenzen und abgeleitete Proteinsequenzen von $\alpha 1$ (IV) und $\alpha 2$ (IV)

Die cDNA-Klone p5 und pFAC (Abb.8, S. 32) überdecken 495 Basen (165 Aminosäuren) der C-terminalen tripelhelikalen Region der $\alpha 1$ -Kette. Abbildung 17 zeigt diese cDNA-Sequenzen und die davon abgeleiteten Proteinsequenzen.

Die Proteindaten aus der Sequenzierung von Bromcyan- und Trypsinpeptiden des terminalen Pepsinfragments P1 von $\alpha 1$ (IV) der Maus (Schuppan et al., 1982) sind in Abbildung 17 durch eine Linie gekennzeichnet. Sie stimmen weitgehend mit den von der cDNA abgeleiteten Proteinsequenzen überein. Lediglich an vier Positionen (40, 61, 77 und 78) zeigen sich Unterschiede (in Abb.17 mit Klammern versehen) zu den durch Edman-Abbau bestimmten Proteinsequenzen, andererseits jedoch Übereinstimmung mit entsprechenden Humansequenzen (Babel und Glanville, 1984). Dies ließ daher auf Sequenzierungsfehler bei der Analyse der Mauspeptide schließen.

Durch die cDNA-Sequenzen konnten zwei Lücken, die auf Proteinebene bestanden, geschlossen werden. Es handelt sich hier um die Positionen 1 bis 15 und 80 bis 140 in Abbildung 17.

Für den gesamten in Abbildung 17 dargestellten tripelhelikalen Bereich der $\alpha 1$ -Kette liegen Humanproteinsequenzen vor (Babel und Glanville, 1984). Auch hier zeigt sich, wie im Bereich der NC1-Domäne, eine ausgeprägte Interspezies-homologie. 92 % der korrespondierenden Positionen bei Mensch und Maus sind von identischen Aminosäuren besetzt. Bei den dreizehn unterschiedlichen Aminosäuren (Abb.17, jeweils in der dritten Zeile) handelt es sich zudem in sieben Fällen um Austausch verwandter Aminosäuren.

1	GCTCCGGGGGTTCCAGGCCCTAAGGGAGACCCAGGATTCCAAGGCATGCCGGGCATTGGC	60
1	A P G V P G P K G D P G F Q G <u>M P G I G</u>	20
61	GGCTCTCCAGGGATCACAGGTTCAAAGGGAGATATGGGACTGCCCGGAGTTCCAGGATTT	120
21	<u>G S P G I T G S K G D M G L P G V P G F</u>	40
	(P)	
121	CAAGGTCAGAAAGGGCTTCTGGTCTGCAGGGAGTGAAGGGAGACCAGGGAGACCAAGGT	180
41	<u>Q G Q K G L P G L Q G V K G D Q G D Q G</u>	60
	P I	
181	GTACCCGGCCCTAAAGGTCTCCAAGGTCCCCTGGGCCCCAGGTCCTTACGATGTCATC	240
61	<u>V P G P K G L Q G P P G P P G P Y D V I</u>	80
	(P) A (D) (G) I	
241	AAAGGAGAACCAGGGCTCCCTGGTCTGAGGGTCCCCCTGGTCTTAAAGGACTCCAGGGA	300
81	K G E P G L P G P E G P P G L K G L Q G	100
	Q	
301	CCACCAGGTCCAAAAGGACAGCAAGGTGTGACAGGCTCAGTGGGCTTGCTGGACCTCCA	360
101	<u>P P G P K G Q Q G V T G S V G L P G P P</u>	120
	L L I	
361	GGTGTCCCTGGGTTCGATGGTGTCTCTGGCCAGAAGGGAGAGACAGGACCCTTTGGACCA	420
121	<u>G V P G F D G A P G Q K G E T G P F G P</u>	140
	I M A	
421	CCTGGTCCAAGAGGGTTTCTGGCCCCGCCGGCCCCGATGGGCTGCCAGGATCCATGGGG	480
141	<u>P G P R G F P G P P G P D G L P G S M G</u>	160
	T	
481	CCCCCAGGTACCCCA	495
161	P P G T P	165

Abb.17: cDNA- und abgeleitete Proteinsequenzen der C-terminalen Region der Tripelhelix von $\alpha 1$ (IV)
 Die auf Proteinebene sequenzierten Peptidabschnitte (Schuppan et al., 1982) von $\alpha 1$ (IV) der Maus sind unterstrichen. Hier auftretende Sequenzunterschiede sind in Klammern angegeben. Die nicht eingeklammerten Aminosäuren der dritten Zeile zeigen Unterschiede in den Humansequenzen von $\alpha 1$ (IV) (Babel und Glanville, 1984).

Bei $\alpha 2(IV)$ konnte durch die Sequenzierung der cDNA-Klone pAIIa und pAIIb (Abb.11, S. 42) ein weitaus größerer Bereich der Tripelhelix aufgeklärt werden als bei $\alpha 1(IV)$, und zwar 511 Aminosäuren der C-terminalen Region. Das entspricht ca. 40 % der gesamten Tripelhelix, die insgesamt ca. 1400 Aminosäuren umfaßt. Wie bereits erwähnt, lagen auf Proteinebene Sequenzen von Bromcyan- und Trypsinfragmenten des C-terminalen Pepsinfragments (P2) der Maus (Schwarz et al., 1986) vor, deren Reihenfolge jedoch aufgrund bestehender Lücken nicht bestimmt werden konnte. Durch die vorliegenden, von cDNA abgeleiteten, Protein-daten konnten diese Fragmente angeordnet werden (Abb.18), wobei keinerlei Diskrepanzen zwischen den beiden Sequenzen zu verzeichnen sind.

Für den tripelhelikalen Bereich der $\alpha 2$ -Kette von Typ IV Kollagen sind bisher keine Humansequenzen bekannt, so daß hier ein Homologievergleich zwischen Mensch und Maus nicht möglich ist.

In einer Publikation von Kurkinen et al. (1985) wird die Sequenz eines 143 Aminosäuren langen Abschnitts aus dem tripelhelikalen Bereich der $\alpha 2$ -Kette (Maus) aufgeführt, der den Positionen 1 bis 140 in Abbildung 18 entspricht - bis auf eine Ausnahme. So wurde an Stelle eines Serinrestes in Position 128 (Abb.18), der bei der Sequenzierung der cDNA und des Proteins gefunden wurde, ein Glycin publiziert. Da dieses Glycin wiederum durch Sequenzierung von cDNA und genomischer DNA bestimmt wurde, können Sequenzierungsartefakte als Ursache für diese Diskrepanz als unwahrscheinlich angesehen werden. Vermutlich ist das Vorliegen eines Polymorphismus innerhalb der Spezies Maus hierfür verantwortlich.

Eine andere Ursache hat wahrscheinlich ein weiterer Fall von fehlender Übereinstimmung von $\alpha 2$ -Maussequenzen. Position 95 der $\alpha 2$ -Kette (Abb.18), auf Proteinebene als Arginin identifiziert, liegt in einem Bereich, in dem die Klone pAIIa und pAIIb überlappen. pAIIa enthält an dieser Position das Codon CGG für

1 CCCCCAGGACCACCTCCTCTCATCTGCGGGGATGAAGGACATCAAGGGAGAAAAGGGA 60
 1 P P G P P P L I L P G M K D I K G E K G 20
 61 GATGAAGGACCAATGGGCTGAAAGGATATCTGGGCTTAAAGGCATCCAAGGAATGCC 120
 21 D E G P M G L K G Y L G L K G I Q G M P 40
 121 GGAGTCCCCGGAGTGTCTGGATTCCCTGGGCTACCTGGAAGGCCTGGCTTCATCAAAGGA 180
 41 G V P G V S G F P G L P G R P G F I K G 60
 181 GTCAAGGGAGACATCGGAGTCCCTGGCACACCAGGCTTGCCGGGATTCCCTGGGGTGTCT 240
 61 V K G D I G V P G T P G L P G F P G V S 80
 241 GGCCCTCCTGGAATTACCGGGTTTCCAGGATTACAGGCAGCCGGGGCGAGAAGGGTACT 300
 81 G P P G I T G F P G F T G S R G E K G T 100
 301 CCAGGAGTAGCAGGAGTTTTGGCGAGACCGGCCCTACTGGGGACTTTGGTGACATTGGG 360
 101 P G V A G V F G E T G P T G D F G D I G 120
 361 GACACTGTGGACTTACCAGGGAGCCAGGCCTGAAGGGGAACGGGGCATCACGGGAATT 420
 121 D T V D L P G S P G L K G E R G I T G I 140
 421 CCAGGTCTCAAGGATTCTTTGGAGAGAAAGGAGCTGCGGGTGACATTGGATTTCCTGGG 480
 141 P G L K G F F G E K G A A G D I G F P G 160
 481 ATAACGGGCATGGCTGGAGCTCAGGGCTCTCCTGGACTTAAAGGACAGACAGGCTTTCCA 540
 161 I T G M A G A Q G S P G L K G Q T G F P 180
 541 GGGCTCACAGGGCTGCAAGGGCCACAGGGAGAGCCTGGACGGATCGGAATACCCGGTGAC 600
 181 G L T G L Q G P Q G E P G R I G I P G D 200
 601 AAGGTGATTTTCGGCTGGCCAGGCGTACCAGGTCTACCAGGTTTTCTGGGATCCGGGGC 660
 201 K G D F G W P G V P G L P G F P G I R G 220
 661 ATCAGCGGATTGCACGGCCTGCCAGGCACCAAAGGCTTCCTGGATCACCAGGTGTGTAT 720
 221 I S G L H G L P G T K G F P G S P G V D 240
 721 GCTCATGGAGACCCAGGGTTCCAGGCCCCACTGGAGACAGGGGTGACCGGGGAGAGGCC 780
 241 A H G D P G F P G P T G D R G D R G E A 260
 781 AACACACTTCAGGCCCCGTGGGAGTTCCAGGGCAGAAAGGGGAGCGGGGAACCCAGGG 840
 261 N T L P G P V G V P G Q K G E R G T P G 280
 841 GAACGTGGCCAGCTGGAAGCCAGGACTTCAGGGTTTCCCTGGTATCTCTCCACCATCC 900
 281 E R G P A G S P G L Q G F P G I S P P S 300
 901 AACATCTCTGGGTACCTGGTGATGTAGGCGCACCAGGAATATTGGCTTGCAAGGCTAC 960
 301 N I S G S P G D V G A P G I F G L Q G Y 320
 961 CAAGGCCCTCCAGGACCACCTGGGCCAAACGCACTTCCTGGAATCAAAGGAGATGAGGGG 1020
 321 Q G P P G P P G P N A L P G I K G D E G 340
 1021 AGTCTGGAGCTGCAGGATTCCTGGACAGAAAGGATGGGTTGGGGACCCAGGGCCCCAG 1080
 341 S S G A A G F P G Q K G W V G D P G P Q 360
 1081 GGCCAGCTGGGTGTACTCGGTCTTCCGGGGGAGAAAGGGCCCAAGGGTGACAAGGATTC 1140
 361 G Q P G V L G L P G E K G P K G E Q G F 380

```

1141 ATGGGCAACACCGGGCCCTCTGGGGCCGTGGGTGACAGAGGCCCAAGGACCCAAAGGC 1200
381 M  G N T G P S G A V G D R G P K G P K G 400

1201 GACCAAGGATTCCCAGGTGCTCCTGGCTCTATGGGGTCCCCAGGAATTCCTGGCATCCCC 1260
401 D  Q G F P G A P G S M G S P G I P G I P 420

1261 CAGAAGATTGCTGTCCAGCCTGGAACGCTGGGTCCCCAGGGCAGGAGAGGCCCTTCCTGGG 1320
421 Q  K I A V Q P G T L G P Q G R R G L P G 440

1321 GCCCTGGGAGAGATAGGGCCGCAGGGCCCTCCTGGAGATCCAGGCTTCCGTGGGGCCCCA 1380
441 A  L G E I G P Q G P P G D P G F R G A P 460

1381 GGCAAGGCTGGGCCTCAGGGCAGAGGTGGCGTGTCTGCTGTTCAGGGTTCCGGGGAGAC 1440
461 G  K A G P Q G R G G V S A V P G F R G A P 480

1441 CAAGGGCCCATGGGACACCGGGTCCAGTTGGCCAGGAAGGGGAGCCAGGCCGTCCAGGG 1500
481 Q  G P M G H Q G P V G Q E G E P G R P G 500

1501 AGCCCAGGCCTGCCAGGGGATGCCAGGCCGGAGT 1533
501 S  P G L P G M P G R S 511

```

Abb. 18: cDNA- und abgeleitete Proteinsequenzen des 511 Aminosäuren langen C-terminalen Bereichs der Tripelhelix von $\alpha 2$ (IV)
Eine Linie kennzeichnet die Abschnitte, für die Proteinsequenzen aus der Analyse entsprechender Peptide vorliegen (Schwarz et al., 1986)

Arginin, Klon pAIIb dagegen eindeutig das Codon CAG für Glutamin. Wahrscheinlich ist dieser eine Basenunterschied die Folge von fehlerhafter cDNA-Synthese. So wird für die AMV Reverse Transkriptase eine Fehlerquote (in vitro) von bis zu eins pro tausend eingebauter Basen angegeben (Gopinathan et al., 1979). Möglich ware auch, daß dem Basenaustausch eine Heterogenität der mRNA-Population zugrunde liegt, die durch fehlerhafte Transkription (Fehlerquote: 10^{-4} bis 10^{-5} , nach Stryer, 1981) verursacht wird.

2.2. Vergleich der Proteinsequenzen von $\alpha 1$ (IV) und $\alpha 2$ (IV)

Durch die Aufklärung der 511 C-terminalen Aminosäuren des tripelhelikalen Bereichs der $\alpha 2$ -Kette war es zum ersten Male möglich, korrespondierende tripelhelikale Abschnitte der $\alpha 1$ - und $\alpha 2$ -Kette zu vergleichen. Die entsprechende 510 Aminosäuren lange Sequenz von $\alpha 1$ (IV) der Maus war in den Positionen 1 bis 9, 80, 82, 85, 122, 224 bis 242 und 316 bis 346 (Abb.19) nicht bekannt, daher wurden an diesen Stellen Humanproteinsequenzen eingesetzt.

Das Hauptinteresse galt zunächst der Frage, ob die für Typ IV Kollagen charakteristischen Unterbrechungen der tripelhelikalen Sequenzen in Anzahl und Lokalisation bei $\alpha 1$ (IV) und $\alpha 2$ (IV) übereinstimmen. Wie Abbildung 19 zeigt, weist dieser Bereich neun Abschnitte (I bis IX) mit Störungen der tripelhelikalen Struktur auf. Bei allen fünf Unterbrechungen der Tripeptidstruktur Gly-X-Y in der $\alpha 1$ -Kette (I, III, IV, VI und VIII) finden sich ähnliche Unterbrechungen auch in der $\alpha 2$ -Kette. In der $\alpha 2$ -Kette treten jedoch vier zusätzliche Unterbrechungen auf, die in der $\alpha 1$ -Kette nicht vorhanden sind (II, V, VII und IX). Auffällig ist, daß die nichttripelhelikalen Sequenzen in der $\alpha 2$ -Kette in der Regel länger sind als in der $\alpha 1$ -Kette. Infolgedessen mußten zwei Lücken in die $\alpha 2$ -Kette (Pos.6 und 127) und drei Lücken in die $\alpha 1$ -Kette (Pos.60, 263 und 300) eingefügt werden, um die zwischen den Unterbrechungen liegenden tripelhelikalen Abschnitte optimal zuordnen zu können. Der Vergleich von $\alpha 1$ (IV) und $\alpha 2$ (IV) hinsichtlich der tripelhelikalen Unterbrechungen zeigt somit, daß weder die Anzahl, noch die Länge, noch die Sequenz der Unterbrechungen konserviert ist.

Die einzelnen nichttripelhelikalen Bereiche unterscheiden sich untereinander hinsichtlich ihrer Länge. Die längste Unterbrechung (I) wird durch acht Aminosäuren der $\alpha 1$ -Kette und elf Aminosäuren der $\alpha 2$ -Kette gebildet. Die kurzen Unterbrechungen lassen sich als einfache Deletionen, Insertionen und Substitutionen

α1 (IV) L P+G K P+G S M D H V D M G S M K+G Q K+G D Q G E K+G Q I G
α2 (IV) 1 P P G P P - P L I L P G M K D I K+G E K+G D E G P M G L K+G 29

30 P T G D K+G S R G D P+G T P+G V P+G K D G Q A G H P+G Q P I I
Y L G L K+G I Q G M P+G V P+G V S G F P+G L P+G R P+G F I K+ 59

60 G P K+G D P G L S G T P+G S K+G L P+G P P G S V G S M G L P+
G V K+G D I G V P+G T P+G L P+G F P+G V S G P P+G I T G F P+ 89

90 G S P+G E K+G V P G I P+G S Q G V P+G S P+G E K+G A K+G E K+
G F T G S R G E K+G T P+G V A G V F G E T G P T G D F G D I 119

120 G Q A G L P+G I G I P+G L P+G D K+G D Q G L A G F P+G S P+G
G D T V D L - P G S P+G L K+G E R G I T G I P+G L K+G F F G 148

149 E K+G E K+G S A G T P+G M P+G S P+G P R G S P+G N I G Y P+G
E K+G A A G D I G F P+G I T G M A G A Q G S P+G L K+G Q T G 178

179 S P+G L P+G D E G D K+G L P+G L D G V P+G V K+G E A G L P+G
F P+G L T G L Q G P Q G E P+G R I G I P+G D K+G D F G W P+G 208

209 T P+G P T G P A G Q K+G E P G S D G I P+G S A G E K+G E P G
V P+G L P+G F P+G I R G I S G L H G L P+G T K+G F P G S P+G 238

239 L P+G R G F P+G F P+G S K+G D K+G S K+G E V - G F P+G L A G
V D A H G D P+G F P+G P T G D R G D R G E A N T L P G P V G 268

269 S P+G I P+G V Q G E Q G F M G P P+G P Q G Q P+G L P+G T P+
V P+G Q K+G E R G T P+G E R G P A G S P+G L Q G F P+G I S P 298

299 G H P V E G P P G D Q G P Q G Q P+G L P+G L P+G P M G P P+G
P S N I S G S P+G D V G A P+G I F G L Q G Y Q G P P+G P P+G 328

329 L P+G I D G V K+G D K+G N P G W P+G A P+G V P+G P K+G D P+G
P N A L P+G I K+G D E G S S G A A G F P G Q K G W V G D P+G 358

359 F Q G M P+G I G G S P+G I T G S K+G D M G L P+G V P+G F Q G
P Q G Q P G V L G L P G E K G P K G E Q G F M G N T G P S G 388

389 Q K+G L P+G L Q G V K+G D Q G D Q G V P+G P K+G L Q G P P+G
A V G D R G P K G P K+G D Q G F P+G A P+G S M G S P+G I P+G 418

419 P P+G P Y D V I K G E P G L P G P E G P P G L K G L Q G P P
I P+G K+I A V Q P G T L G P Q G R R G L P+G A L G E I G P Q 448

	G	P	K	G	Q	Q	G	V	T	G	S	V	G	L	P	G	P	P	G	V	P	G	F	D	IX G	A	P	G	Q	K					
449	G	P	P	+	G	D	P	+	G	F	R	G	A	P	+	G	K	A	G	P	Q	G	R	G	V	S	A	V	P	+	G	F	R	478	
	G	E	T	G	P	F	G	P	P	G	P	R	G	F	P	+	G	P	P	+	G	P	D	G	L	P	+	G	S	M	G	P	P	+	
479	G	D	Q	G	P	M	G	H	Q	G	P	V	G	Q	E	G	E	P	+	G	R	P	+	G	S	P	+	G	L	P	G	M	P	+	508
	G	T	P																																
509	G	R	S	511																															

Abb. 12: Parallele Anordnung der tripelhelikalen Peptidsequenzen von $\alpha 1(IV)$ und $\alpha(IV)$

Die in diesem Bereich auftretenden Unterbrechungen der tripelhelikalen Sequenzen sind hervorgehoben und nummeriert (I bis IX). Hydroxylierte Prolin- und Lysinreste sind mit einem "+" versehen.

interpretieren. In zwei Fällen (VII und IX) wurde jeweils ein Glycin gegen ein Alanin ausgetauscht. In einem anderen Fall (II) ist ein Lysinrest zwischen zwei Triplets eingeschoben. Unterbrechung IV wurde durch je eine Deletion in $\alpha 1(IV)$ und $\alpha 2(IV)$ verursacht. Die in der Länge variablen nichttripelhelikalen Abschnitte wirken auch in unterschiedlicher Weise auf die Struktur der Tripelhelix. Hierauf wird im Diskussionsteil eingegangen.

Die nichttripelhelikale Region VI enthält in der Sequenz Asn-Ile-Ser eine potentielle Carbohydratanheftungsstelle für Oligosaccharide, die über die Amino- gruppe von Asparagin gebunden werden. Für eine Glykosylierung des betreffenden Asparagins spricht, daß diese Position in der Peptidsequenzanalyse durch Edman- Abbau nicht nachzuweisen war.

Nachdem die Analyse der tripelhelikalen Unterbrechungen gezeigt hatte, daß hier nur bedingt Homologien zwischen $\alpha 1(IV)$ und $\alpha 2(IV)$ bestehen, wurden die beiden α -Peptidketten insgesamt auf Sequenzhomologien hin untersucht. Um die Ähnlichkeit beider Abschnitte zunächst optisch darzustellen, wurden alle identischen und verwandten Aminosäuren - Glycin in jeder dritten Position ausgenommen - eingerahmt (Abb.20). Hier zeigt sich, daß nur relativ wenige Reste in den X- und Y-Positionen der Tripeptide (Gly-X-Y) in beiden Ketten homolog sind. 75 Aminosäuren sind identisch, wobei es sich um isolierte Positionen handelt. Es treten also keine Blöcke identischer Aminosäuren auf wie im Bereich der NC1-Domäne. 61 Positionen werden von verwandten Aminosäuren besetzt. Auffällig ist, daß 60 % der identischen Aminosäuren durch Prolinreste gestellt werden, und zwar meistens in der Y-Position der Triplets (Gly-X-Y). Auch der Prozentsatz an Lysinen unter den identischen Aminosäuren ist vergleichsweise hoch (11 %) und betrifft ausschließlich die Y-Position. Diese Befunde sind insofern bemerkenswert, als bei Typ IV Kollagen bis auf wenige Ausnahmen alle Proline und Lysine in der Y-Position hydroxyliert und die Hydroxylysine außerdem glykosyliert sind (Schuppan et al., 1982; Babel und Glanville, 1984; Schwarz et al., 1986). Offensichtlich wurden die Positionen eines Großteils dieser posttranslational modifizierten Aminosäuren in beiden Ketten konserviert. Die $\alpha 1(IV)$ -Kette enthält mehr Lysine als die $\alpha 2(IV)$ -Kette (31 gegenüber 19). Dieses Defizit wird in der $\alpha 2$ -Kette durch einen höheren Gehalt an der ebenfalls basischen Aminosäure Arginin kompensiert, die auch häufig die gleichen Positionen wie Lysin in der $\alpha 1$ -Kette besetzt. Ein weiterer hoher Anteil der homologen Positionen kommt durch Austausch von Prolin, Serin und Alanin zueinander zustande.

$\alpha 1$ (IV)

$\alpha 2$ (IV)

	L	P	P	K	P	G	S	M	D	H	V	D	M	G	S	M	K	G	Q	K	G	D	Q	G	E	K	G	Q	I	G		
1	P	P	G	P	P	-	P	L	I	L	P	G	M	K	D	I	K	G	E	K	G	D	E	G	P	M	G	L	K	G	29	
30	P	T	G	D	K	G	S	R	G	D	P	G	T	P	G	V	P	G	K	D	G	Q	A	G	H	P	G	Q	P	-		
	Y	L	G	L	K	G	I	Q	G	M	P	G	V	P	G	V	S	G	F	P	G	L	P	G	R	P	G	F	I	K	+	59
60	G	P	K	G	D	P	G	L	S	G	T	P	G	S	K	G	L	P	G	P	P	G	S	V	G	S	M	G	L	P	+	
	G	V	K	G	D	I	G	V	P	G	T	P	G	L	P	G	F	P	G	V	S	G	P	P	G	I	T	G	F	P	+	89
90	G	S	P	G	E	K	G	V	P	G	I	P	G	S	Q	G	V	P	G	S	P	G	E	K	G	A	K	G	E	K	+	
	G	F	T	G	S	R	G	E	K	G	T	P	G	V	A	G	V	F	G	E	T	G	P	T	G	D	F	G	D	I	119	
120	G	Q	A	G	L	P	G	I	G	I	P	G	L	P	G	D	K	G	D	Q	G	L	A	G	F	P	G	S	P	G		
	G	D	T	V	D	L	-	P	G	S	P	G	L	K	G	E	R	G	I	T	G	I	P	G	L	K	G	F	F	G	148	
149	E	K	G	E	K	G	S	A	G	T	P	G	M	P	G	S	P	G	P	R	G	S	P	G	N	I	G	Y	P	G		
	E	K	G	A	A	G	D	I	G	F	P	G	I	T	G	M	A	G	A	Q	G	S	P	G	L	K	G	Q	T	G	178	
179	S	P	G	L	P	G	D	E	G	D	K	G	L	P	G	L	D	G	V	P	G	V	K	G	E	A	G	L	P	G		
	F	P	G	L	T	G	L	Q	G	P	Q	G	E	P	G	R	I	G	I	P	G	D	K	G	D	F	G	W	P	G	208	
209	T	P	G	P	T	G	P	A	G	Q	K	G	E	P	G	S	D	G	I	P	G	S	A	G	E	K	G	E	P	G		
	V	P	G	L	P	G	F	P	G	I	R	G	I	S	G	L	H	G	L	P	G	T	K	G	F	P	G	S	P	G	238	
239	L	P	G	R	G	F	P	G	F	P	G	S	K	G	D	K	G	S	K	G	E	V	-	G	F	P	G	L	A	G		
	V	D	A	H	G	D	P	G	F	P	G	P	T	G	D	R	G	D	R	G	E	A	N	T	L	P	G	P	V	G	268	
269	S	P	G	I	P	G	V	Q	G	E	Q	G	F	M	G	P	P	G	P	Q	G	Q	P	G	L	P	G	T	P	+		
	V	P	G	Q	K	G	E	R	G	T	P	G	E	R	G	P	A	G	S	P	G	L	Q	G	F	P	G	I	S	P	298	
299	G	H	P	V	E	G	P	P	G	D	Q	G	P	Q	G	Q	P	G	L	P	G	L	P	G	P	M	G	P	P	G		
	P	S	N	I	S	G	S	P	G	D	V	G	A	P	G	I	F	G	L	Q	G	Y	Q	G	P	P	G	P	P	G	328	
329	L	P	G	I	D	G	V	K	G	D	K	G	N	P	G	W	P	G	A	P	G	V	P	G	P	K	G	D	P	G		
	P	N	A	L	P	G	I	K	G	D	E	G	S	S	G	A	A	G	F	P	G	Q	K	G	W	V	G	D	P	G	358	
359	F	Q	G	M	P	G	I	G	G	S	P	G	I	T	G	S	K	G	D	M	G	L	P	G	V	P	G	F	Q	G		
	P	Q	G	Q	P	G	V	L	G	L	P	G	E	K	G	P	K	G	E	Q	G	F	M	G	N	T	G	P	S	G	388	
389	Q	K	G	L	P	G	L	Q	G	V	K	G	D	Q	G	D	Q	G	V	P	G	S	K	G	L	Q	G	P	P	G		
	A	V	G	D	R	G	P	K	G	P	K	G	D	Q	G	F	P	G	A	P	G	S	M	G	S	P	G	I	P	G	418	
419	P	P	G	P	Y	D	V	I	K	G	E	P	G	L	P	G	P	E	G	P	P	G	L	K	G	L	Q	G	P	P		
	I	P	Q	K	I	A	V	Q	P	G	T	L	G	P	Q	G	R	R	G	L	P	G	A	L	G	E	I	G	P	Q	448	
449	G	P	K	G	Q	Q	G	V	T	G	S	V	G	L	P	G	P	P	G	V	P	G	F	D	G	A	P	G	Q	K		
	G	P	P	G	D	P	G	F	R	G	A	P	G	K	A	G	P	Q	G	R	G	G	V	S	A	V	P	G	F	R	478	
479	G	E	T	G	P	F	G	P	P	G	P	R	G	F	P	G	P	P	G	P	D	G	L	P	G	S	M	G	P	P		
	G	D	Q	G	P	M	G	H	Q	G	P	V	G	Q	E	G	E	P	G	R	P	G	S	P	G	L	P	G	M	P	508	
509	G	T	P																													
	G	R	S																													

Abb. 20: Parallele Anordnung der tripelhelikalen Peptidsequenzen von $\alpha 1$ (IV) und $\alpha 2$ (IV)
Identische Aminosäuren sind durch eine Linie, verwandte durch Punkte gekennzeichnet. Hydroxylierte Prolin- und Lysinreste sind an einem "+" erkennbar.

Die Verteilung einzelner Aminosäuren auf X- oder Y-Position der Triplets (Gly-X-Y) ist sehr ähnlich bei $\alpha 1(IV)$ und $\alpha 2(IV)$ (Tab.6) und zeigt weitgehend die gleichen Charakteristiken wie bei den interstitiellen Kollagenen. Lediglich die Überrepräsentation von Prolin und Lysin in der Y-Position unterscheidet Typ IV Kollagen von den interstitiellen Kollagenen, wo diese Aminosäuren gleichmäßig zwischen X- und Y-Position verteilt sind. Ansonsten ist auch bei Typ IV Kollagen eine Bevorzugung der aromatischen Aminosäuren für die X-Position zu beobachten, gleiches gilt für Glutaminsäure und Leucin. Arginin da-

Tab.6: Verteilung der Aminosäuren zwischen X- und Y-Position

Aminosäure	$\alpha 1(IV)$		$\alpha 2(IV)$	
	X	Y	X	Y
Asp	15	6	17	10
Asn	2	0	1	1
Thr	5	5	5	10
Ser	22	1	10	7
Glu	14	2	13	3
Gln	12	15	5	15
Pro	28	78	23	55
Gly	0	1	0	1
Ala	3	8	8	7
Val	12	3	12	5
Met	2	6	2	4
Ile	9	2	11	6
Leu	24	0	17	4
Tyr	1	0	2	0
Phe	9	1	17	5
His	1	0	1	1
Lys	2	29	1	18
Arg	0	3	5	11
Trp	1	0	2	0

gegen tritt vorwiegend in der Y-Position auf. Nur die Verteilung von Serin in der $\alpha 1(IV)$ -Kette ist ungewöhnlich. Serin findet sich hier fast ausschließlich in der X-Position, im Unterschied zur $\alpha 2(IV)$ -Kette und auch zu den interstitiellen Kollagenen.

Trotz der schwach ausgeprägten Homologie hinsichtlich des Auftretens identischer und verwandter Aminosäuren an gleichen Positionen gibt es kaum Differenzen in der Aminosäurezusammensetzung der beiden Ketten, wenn diese nach dem Gehalt verwandter Aminosäuren aufgegliedert wird (Tab.7). Lediglich der Prozentsatz an den aromatischen Aminosäuren ist in der $\alpha 2$ -Kette mehr als doppelt so hoch wie in der $\alpha 1$ -Kette, was vor allem durch den höheren Anteil an Phenylalanin (22 gegenüber 10) verursacht wird.

Tab.7. Gehalt homologer Aminosäuren im tripelhelikalen Bereich von $\alpha 1(IV)$ und $\alpha 2(IV)$, Glycin ausgenommen

Aminosäuren	$\alpha 1(IV)$		$\alpha 2(IV)$	
	Anzahl	Prozentsatz	Anzahl	Prozentsatz
P, A, S, T	152	29.95 %	143	28.1 %
Q, N, E, D	65	12.7 %	63	12.4 %
H, K, R	42	8.3 %	42	8.3 %
L, I, V, M	69	13.55 %	77	15.1 %
F, Y, W	12	2.4 %	26	5.1 %

Zur Quantifizierung des Homologiegrades der tripelhelikalen Abschnitte von $\alpha 1(IV)$ und $\alpha 2(IV)$ wurde eine Computeranalyse durchgeführt. Hierzu mußten die tripelhelikalen Unterbrechungen herausgenommen werden. Das Computerprogramm "Relate" (Barker et al., 1983) bewertet neben dem Auftreten identischer Positionen auch die Austausche verwandter Aminosäuren, wobei übereinstimmende Positionen seltener Aminosäuren höher eingestuft werden als die häufiger Aminosäuren. Der Grad der Homologie wird in S.D. Einheiten angegeben. Werte unter 3 S.D. besagen, daß keinerlei Homologie vorhanden ist. Die Anwendung dieses "Relate"-Programmes ergab für die Homologie des tripelhelikalen Bereichs von $\alpha 1(IV)$ und $\alpha 2(IV)$ einen Wert von 3.7 S.D. Einheiten, also ein ausgesprochen niedriger Homologiegrad. Das Auftreten von Glycin in jeder dritten Position wurde hierbei nicht als Ausdruck von Homologie gewertet.

Um festzustellen, ob der tripelhelikale Bereich von Typ IV Kollagen Sequenzhomologien zu den in der Primärstruktur untereinander sehr ähnlichen interstitiellen Kollagenen aufweist, wurde $\alpha 1(IV)$ mit $\alpha 1(I)$ (Huhn) verglichen. Es zeigte sich, daß zwischen den beiden Ketten keine Sequenzhomologien ausgeprägt sind.

Ein Vergleich der α -Ketten $\alpha 1(I)$ und $\alpha 2(I)$ von Typ I Kollagen (Huhn) untereinander dagegen lieferte einen Wert von 16 S.D., was ein hohes Maß an Verwandtschaft bedeutet. Die Gegenüberstellung der beiden stark differierenden Homologiegrade von Typ I (16 S.D.) und Typ IV (3.7 S.D.) Kollagen verdeutlicht, daß die tripelhelikalen Sequenzen der beiden Typ IV α -Ketten kaum konserviert sind, und daß diese Eigenschaft Typ IV Kollagen von den interstitiellen Kollagenen unterscheidet. Diese Befunde stehen im Zusammenhang mit der besonderen makromolekularen Struktur, die Typ IV Kollagen im Gegensatz zu den interstitiellen Kollagenen ausbildet, worauf im Diskussionsteil eingegangen werden wird.

3. Analyse der codierenden cDNA-Sequenzen von $\alpha 1$ (IV) und $\alpha 2$ (IV)

Die Vergleiche der Peptidsequenzen der $\alpha 1$ - und $\alpha 2$ -Kette von Typ IV Kollagen hatten gezeigt, daß die tripelhelikalen Bereiche beider Ketten kaum Homologien aufweisen, während die NC1-Sequenzen am C-Terminus in hohem Ausmaß konserviert sind. Nun war von Interesse, inwieweit auf DNA-Ebene in den entsprechenden Genabschnitten von $\alpha 1$ (IV) und $\alpha 2$ (IV) Homologien ausgeprägt sind. Auch die bereits bekannten cDNA-Sequenzen der humanen $\alpha 1$ (IV)-Kette (Pihlajaniemi et al., 1985) wurden in die Homologievergleiche miteinbezogen.

Außerdem wurden die einzelnen Genabschnitte auf ihre jeweilige Codonbenutzung und ihren Nukleotidgehalt hin analysiert. Diese Strukturmerkmale konnten dann wiederum bei $\alpha 1$ (IV) (Mensch und Maus) und $\alpha 2$ (IV) (Maus) verglichen, sowie entsprechenden Daten anderer Kollagen-Gene gegenübergestellt werden.

3.1. Sequenzhomologien der cDNA von $\alpha 1$ (IV) und $\alpha 2$ (IV)

Zunächst wurde untersucht, ob sich die auf Proteinebene beobachteten mehr oder weniger stark ausgeprägten Homologien von $\alpha 1$ (IV) und $\alpha 2$ (IV) in ähnlichem Ausmaß auf DNA-Ebene manifestieren. Die Frage war, ob die in beiden Ketten identischen Aminosäurepositionen auch durch identische Codons bestimmt werden. In Tab.8 sind Anzahl und prozentualer Anteil von übereinstimmenden Codons für identische Aminosäuren aufgelistet, und zwar getrennt für Tripelhelix und NC1-Domane. Bei der Tripelhelix wurde jeweils der 165 Aminosäuren (495 bp) lange C-terminale Abschnitt analysiert. In der Tripelhelix ist der überwiegende Teil der identischen Positionen von Glycinen besetzt. Daher sind die entsprechenden Werte für diese Aminosäure zusätzlich getrennt angegeben.

Tab.8: übereinstimmende Codons für identische Aminosäurepositionen im Bereich der Tripelhelix und der NC1-Domäne von $\alpha 1(IV)$ und $\alpha 2(IV)$ von Maus, sowie von $\alpha 1(IV)$ von Maus und Mensch (Pihlajaniemi et al., 1985)

	identische Aminosäuren	übereinstimmende Codons für identische Aminosäuren
$\alpha 1 : \alpha 2$ (Maus)		
Tripelhelix (165 Positionen)	75 (davon 51 Gly)	26 (35 %) (davon 16 für Gly)
NC1-Domäne (230 Positionen)	145	68 (47 %)
$\alpha 1$ (Mensch) : $\alpha 1$ (Maus)		
Tripelhelix (165 Positionen)	152 (davon 53 Gly)	89 (59 %) (davon 27 für Gly)
NC1-Domäne (229 Positionen)	222	148 (67 %)

Die Analyse zeigt, daß im Bereich der NC1-Domäne (Maus) nur 47 % der bei $\alpha 1(IV)$ und $\alpha 2(IV)$ übereinstimmenden Aminosäurepositionen auch von identischen Codons bestimmt werden. Es unterscheiden sich also rund die Hälfte der betreffenden Codons hinsichtlich einer Base in der dritten Position, bzw. bei Leucin, Serin und Arginin (je sechs Codons) zum Teil auch in der ersten und zweiten Position. Dieser Prozentsatz von 47 % liegt nur knapp über dem Wert, der bei einer rein statistischen Verteilung der Codons zu erwarten wäre. So läßt sich - in Abhängigkeit von der jeweiligen Codonbenutzung von $\alpha 1(IV)$ und $\alpha 2(IV)$ im NC1-Bereich (Tab.10) - die Wahrscheinlichkeit für das Auftreten übereinstimmender

Codons für identische Aminosäuren berechnen. Der entsprechende Prozentsatz liegt bei ca. 42 %. Durch diesen statistischen Wert erhält man ein Maß dafür, wie die tatsächliche Übereinstimmung, nämlich 47 %, zu bewerten ist. Die Differenz von 5 % ist zwar signifikant, bedeutet aber, daß die DNA-Sequenzen des NC1-Bereichs von $\alpha 1(IV)$ und $\alpha 2(IV)$ kaum konserviert sind.

In der Tripelhelix ist der Anteil identischer Codons für identische Aminosäuren noch niedriger als im Bereich der NC1-Domäne - hier ist nur in 35 % der Fälle eine Übereinstimmung vorhanden (Tab.8). Bei diesem Wert muß allerdings berücksichtigt werden, daß der größte Anteil identischer Aminosäuren in der Tripelhelix durch Glycine gestellt wird (51 von 75) und auch Prolin sehr häufig ist (17 von 75). Nachdem Glycin und Prolin jeweils durch vier Codons bestimmt werden, kann es hier zu vergleichsweise überdurchschnittlich vielen Basenaustauschen in der dritten Position der Basentriplets kommen, ohne daß die codierte Position im Peptid verändert wird. So wäre für den tripelhelikalen Bereich bei einer rein statistischen Verteilung der Codons ein Prozentsatz von ca. 28 % Übereinstimmung der Codons für identische Aminosäuren zu erwarten. Der tatsächliche Wert von 35 % liegt zwar noch signifikant höher, spiegelt aber - wie auch im NC1-Bereich - nur einen niedrigen Grad an Homologie wieder.

Die mangelnde Konservierung der Codons für identische Aminosäuren bei $\alpha 1(IV)$ und $\alpha 2(IV)$ - sowohl in der Tripelhelix als auch in der NC1-Domäne - läßt sich dahingehend interpretieren, daß der Evolutionsdruck selektiv auf die Konservierung der Primärstruktur der Peptide gerichtet ist und nicht auf die der codierenden DNA-Sequenzen. Möglicherweise stellt die Divergenz der DNA-Sequenzen auch einen Schutzmechanismus dar, der die Wahrscheinlichkeit von Rekombinationen der Gene von $\alpha 1(IV)$ und $\alpha 2(IV)$ reduziert.

Der Vergleich der codierenden DNA-Abschnitte von $\alpha 1(IV)$ der Maus mit entsprechenden $\alpha 1(IV)$ -Humansequenzen (Pihlajaniemi et al., 1985) zeigt (Tab.8), daß hier Homologien hinsichtlich identischer Codons stärker ausgeprägt sind als beim Vergleich der beiden α -Ketten von Maus untereinander. Die berechneten Anteile von 67 % im NC1-Bereich und 59 % in der Tripelhelix liegen deutlich höher als die entsprechenden Werte (47 % und 35 %) bei der Gegenüberstellung von $\alpha 1(IV)$ und $\alpha 2(IV)$.

Aus diesem Befund kann man ableiten, daß der Verwandtschaftsgrad der $\alpha 1(IV)$ -Gene von Mensch und Maus weitaus stärker entwickelt ist als der der beiden Typ IV Kollagengene innerhalb der Spezies Maus. Dies wiederum weist darauf hin, daß die Divergenz der Mensch- und Mausgene in der Evolution zu einem späteren Zeitpunkt stattgefunden hat als die Entstehung von $\alpha 1(IV)$ und $\alpha 2(IV)$ aus einem postulierten gemeinsamen Vorläufergen.

3.2. Nukleotidzusammensetzung und Codonauswahl

Die Codonauswahl von Genen erfolgt nicht zufällig und ist ein charakteristisches Strukturmerkmal eines Gens. Bei taxonomisch verwandten Organismen ist eine ähnliche Tendenz der Codonpräferenzen festzustellen (Grantham et al., 1981), jedoch kann innerhalb einer Spezies die Codonauswahl von Gen zu Gen stark variieren. Über das Prinzip, daß der Ausprägung eines spezifischen Musters der Codonenbenutzung zugrunde liegt, ist, insbesondere bei höheren Eukaryonten, kaum etwas bekannt. Man geht auch davon aus, daß mehrere Faktoren eine Rolle spielen.

Bei Bakterien und auch bei der Hefe hat man eine Korrelation zwischen der Expressionshäufigkeit von Genen und der Codonauswahl gefunden (Ikemura, 1981; Bennetzen und Hall, 1982; Grosjean und Fears, 1982; Sharp et al., 1986). Bei stark exprimierten Genen sind die Codonpräferenzen offensichtlich an das Auftreten häufiger tRNAs angepaßt. Außerdem ist hier die Codonauswahl oft daraufhin aus-

gerichtet, daß die Wechselwirkung der Codon/Anticodon-Paarung optimal ist (nicht zu stark und nicht zu schwach), so daß die Effizienz der Translation möglichst hoch ist. Für die optimalen Interaktionen von Codon und Anticodon sollte der GC-Gehalt des jeweiligen Gens bzw. seiner mRNA eine wichtige Rolle spielen. Bei Bakterien und der Hefe hat man auch eine Korrelation zwischen dem GC-Gehalt und der Codonauswahl gefunden. So erscheint es auch sinnvoll, daß AT-reiche Sequenzen durch einen hohen GC-Gehalt in den "wobble"-Positionen die Stabilität von Wechselwirkungen erhöhen, während GC-reiche Sequenzen bevorzugt A und T in den variablen Positionen der Codons enthalten. Nachdem der GC-Gehalt generell bei der Ausbildung und der Stabilität von Sekundärstrukturen von Nukleinsäuren eine Rolle spielt, besteht die Möglichkeit, daß die jeweilige Codonauswahl hier eine wichtige Komponente darstellt.

Die codierenden Sequenzen der Kollagene haben im tripelhelikalen Bereich erwartungsgemäß einen sehr hohen GC-Gehalt (Tab.9), verursacht durch das häufige Auftreten der Codons für Glycin (GGX), Prolin (CCX) und bei den Kollagenen Typ I-III auch für Alanin (GCX). Im Vergleich dazu ist der GC-Gehalt der C-terminalen Sequenzen (Carboxypropeptide und NC1-Domäne) generell reduziert (Tab.9). Bei der Analyse der Basenpräferenzen für die dritten Positionen der Codons in den tripelhelikalen und C-terminalen Bereichen wird deutlich, daß die Nukleotidzusammensetzung der "wobble"-Position durch den Gesamt-GC-Gehalt des jeweiligen Genabschnitts (Tripelhelix und C-terminale Domäne) beeinflusst wird. In GC-reichen tripelhelikalen Bereichen ist der GC-Anteil in den "wobble"-Positionen deutlich verringert - offensichtlich ist ein Selektionsdruck auf die Verwendung von Codons mit einem A oder T in der dritten Position gerichtet. Dadurch wird der hohe Gesamt-GC-Gehalt bis zu einem gewissen Grad kompensiert. Die gegenläufige Tendenz ist in den Genabschnitten zu beobachten, die für die C-terminalen Domänen (Carboxypropeptide und NC1-Domänen) der Kollagene codieren. Dem insgesamt niedrigerem Gesamt-GC-Gehalt stehen hier in den meisten Fällen deutlich höhere GC-Anteile der "wobble"-Position gegenüber.

Nukleotid- zusammensetzung	Nukleotid	TRIPSELHELIX					
		$\alpha 1(I)^a$ (Huhn)	$\alpha 2(I)^a$ (Huhn)	$\alpha 1(I)^b$ (Maus)	$\alpha 1(IV)^d$ (Mensch)	$\alpha 1(IV)$ (Maus)	$\alpha 2(IV)$ (Maus)
Gesamt	T	14.4	26.0	23.1	18.6	18.0	16.9
	C	41.5	28.2	31.3	26.3	28.5	27.2
	A	10.4	14.1	11.1	18.7	20.9	21.2
	G	33.7	31.7	34.4	36.4	32.6	35.0
	GC/AT	3.0	1.50	1.92	1.68	1.57	1.67
in der "wobble"- Position	wobble						
	T	35.3	65.7	61.9	36.6	30.8	26.5
	C	48.2	14.9	25.4	27.1	21.6	21.0
	A	9.5	12.4	7.6	18.3	28.6	30.9
	G	7.0	7.0	5.1	18.0	18.9	21.6
	GC/AT	1.22	0.28	0.44	0.82	0.68	0.74
analysierte Codons		201	201	180	469	185	162

Nukleotid- zusammensetzung	Nukleotid	CARBOXYPROPEPTID				NC1-DOMÄNE	
		$\alpha 1(I)^a$ (Huhn)	$\alpha 2(I)^a$ (Huhn)	$\alpha 1(III)^e$ (Huhn)	$\alpha 1(IV)^d$ (Mensch)	$\alpha 1(IV)$ (Maus)	$\alpha 2(IV)$ (Maus)
Gesamt	T	14.5	23.1	24.4	21.8	21.1	19.6
	C	32.4	24.9	20.3	29.0	30.9	31.4
	A	24.2	29.0	29.7	22.2	22.4	21.1
	G	28.6	23.1	25.6	21.8	25.6	27.9
	GC/AT	1.57	0.92	0.85	1.26	1.30	1.46
in der "wobble"- Position	wobble						
	T	8.1	30.8	33.6	18.3	16.6	14.0
	C	55.9	35.0	25.4	38.9	43.7	45.2
	A	5.5	17.3	21.6	13.5	13.5	9.6
	G	30.5	16.9	19.4	29.3	26.2	31.1
	GC/AT	6.35	1.1	0.81	2.1	2.3	3.2
analysierte Codons		272	260	268	229	229	228

- ^a Fuller und Boedtker, 1981
^b Monson und McCarthy, 1981
^c Monson et al., 1982
^d Pihlajaniemi et al., 1985
^e Yamada et al., 1983

Tab.9: Nukleotidzusammensetzung und Basenpräferenzen für die "wobble"-Positionen der Codons bei verschiedenen Kollagentypen
Die entsprechenden Werte sind für die tripelhelikalen Bereiche und die C-terminalen Domänen (Carboxypropeptide und NC1-Region) getrennt aufgelistet.

Trotzdem sich der GC-Gehalt bei den verschiedenen Kollagenen in ähnlicher Weise auf die Codonauswahl auswirkt, zeigen sich beim Vergleich der Kollagene untereinander zum Teil große Unterschiede in der Nukleotidzusammensetzung und den Basenpräferenzen. Bei den interstitiellen Kollagenen sind im tripelhelikalen Bereich die "wobble"-Positionen der Codons vor allem durch das Nukleotid T besetzt (bis zu 65 %, Tab.9). Bei den Typ IV-Genen dagegen sind A und T in den variablen Positionen der Tripelhelix annähernd gleich stark vertreten; bei der Maus überwiegt A leicht, bei $\alpha 1(IV)$ von Mensch dagegen T. Unter den Typ IV-Genen zeichnet sich $\alpha 2(IV)$ durch das höchste GC/AT-Verhältnis in den "wobble"-Positionen des tripelhelikalen Bereichs aus, $\alpha 1(IV)$ von Mensch zeigt das niedrigste. In ihrer gesamten Nukleotidzusammensetzung innerhalb der tripelhelikalen Region stimmen $\alpha 1(IV)$ und $\alpha 2(IV)$ von Maus weitgehend überein, $\alpha 1(IV)$ von Mensch hat einen etwas geringeren GC-Gehalt.

Hinsichtlich der Nukleotidzusammensetzung im tripelhelikalen Bereich ähneln die Typ IV Gene noch am ehesten dem Drosophila Kollagen (Monson et al., 1982), das auch ein dem Basalmembrankollagen ähnliches Protein darstellt. Die Nukleotidzusammensetzungen der Gene interstitieller Kollagene differieren zum Teil deutlich von denen der Typ IV Gene, unterscheiden sich aber auch untereinander. Besonders auffällig wird dies beim Vergleich der Gene von Typ I Kollagen von Huhn und Maus. So zeigen sich bei $\alpha 1(I)$ von Huhn (Fuller und Boedtke, 1981) im Vergleich zu $\alpha 2(I)$ von Huhn (Fuller und Boedtke, 1981) und $\alpha 1(I)$ von Maus (Monson und McCarthy, 1981) ungewöhnlich hohe Werte sowohl für den Gesamt-GC-Gehalt als auch für die Basenpräferenzen in der "wobble"-Position. Entsprechend differiert die Codonauswahl sehr stark. Dieser Befund verdeutlicht, wie sehr die Codonauswahl nicht nur zwischen zwei homologen Genen ein und desselben Organismus ($\alpha 1(I)$ und $\alpha 2(I)$ von Huhn) variieren kann, sondern auch zwischen zwei homologen Genen verschiedener Spezies ($\alpha 1(I)$ von Huhn und Maus). Da $\alpha 1(I)$ und

$\alpha 2(I)$ zumindest beim Menschen auf verschiedenen Chromosomen liegen (Huerre et al., 1982; Solomon et al., 1983), zieht man die Möglichkeit in Betracht, daß die jeweilige Codonauswahl durch die chromosomale Lokalisation beeinflusst wird (Boedtker et al., 1983).

Im NC1-Bereich stimmen die Nukleotidzusammensetzungen von $\alpha 1(IV)$ und $\alpha 2(IV)$ von Maus ebenfalls sehr gut überein (Tab.9). Auch hier ist, wie in der tripelhelikalen Region, die Tendenz festzustellen, daß $\alpha 1(IV)$ von Mensch einen etwas niedrigeren GC-Gehalt aufweist (auch in der "wobble"-Position), während $\alpha 2(IV)$ im NC1-Bereich vor allem in der dritten Position der Basentriplets einen vergleichsweise hohen GC-Anteil zeigt.

Unter den Sequenzen der Carboxypropeptide vom C-Terminus der interstitiellen Kollagene zeichnet sich abermals $\alpha 1(I)$ von Huhn durch den höchsten GC-Gehalt aus. $\alpha 2(I)$ und $\alpha 1(III)$ von Huhn liegen deutlich niedriger und erreichen auch nicht den GC-Gehalt der Typ IV-Gene.

Die Codonauswahl für die einzelnen Aminosäuren der Typ IV Kollagenketten sind in Tab.10 aufgeführt, und zwar getrennt für den tripelhelikalen und den NC1-Bereich. Hier wird deutlich, daß die Codonpräferenzen nicht nur von Gen zu Gen variieren (mehr oder weniger), sondern auch in zwei Abschnitten ein und des selben Gens (Tripelhelix und NC1-Domäne) unterschiedlich gelagert sein können. Dies war auch aufgrund der unterschiedlichen Basenauswahl für die "wobble"-Positionen im tripelhelikalen und im NC1-Bereich (Tab.9) zu erwarten. So zeigt sich z.B. bei $\alpha 1(IV)$ von Maus, daß im tripelhelikalen Abschnitt für Glycin und Prolin die Nukleotide A und T in der dritten Position der Codons bevorzugt werden, im Gegensatz zum NC1-Bereich, wo G und C deutlich überwiegen.

Vergleicht man die Codonverwendung (Tab.10) von $\alpha 1(IV)$ und $\alpha 2(IV)$ insgesamt (Tripelhelix und NC1-Domäne), sind hier zum Teil sehr gute Übereinstimmungen (z.B. für Glu und Lys) festzustellen, andererseits aber auch deutliche Differenzen (z.B. für Arg und Val). Bezieht man $\alpha 1(IV)$ von Mensch in den Vergleich mit ein, so zeichnet sich die Tendenz ab, daß sich die $\alpha 1(IV)$ -Gene von Mensch und Maus in ihrer Codonbenutzung ähnlicher sind als $\alpha 1(IV)$ und $\alpha 2(IV)$ innerhalb der Spezies Maus (z.B. für Glu, Arg, Thr, Cys und His).

Bei der Gegenüberstellung des Codongebrauchs der Typ IV-Gene mit dem der Gene interstitieller Kollagene fällt vor allem die unterschiedliche Verwendung der Codons für die Aminosäuren Glycin und Prolin insbesondere im tripelhelikalen Abschnitt auf. Bei den interstitiellen Kollagenen codieren fast ausschließlich Basentriplets mit einem C oder T in der dritten Position (Boedtker et al., 1983), nämlich in der Regel über 80 % und bis zu 100 %. Der entsprechende Anteil bei Typ IV Kollagen liegt durchschnittlich bei nur ca. 50 % (Tab.10). Außerdem tritt bei den interstitiellen Kollagenen das Codon GGG für Glycin fast gar nicht auf, bei den Typ IV Genen dagegen kommt es sogar relativ häufig vor (17 % bis 35 %), und zwar sowohl in der tripelhelikalen, wie auch in der NC1-Region.

Die Ursachen für zum Teil starken Differenzen des Codongebrauchs von strukturell und genetisch verwandten Proteinen wie den Kollagenen sind, wie bereits erwähnt, unklar. Der offensichtliche Einfluß des GC-Gehalts eines Gens bzw. eines Genabschnitts auf die Basenpräferenzen für die "wobble"-Position wurde jedoch beim Vergleich Tripelhelix - C-terminale Domäne deutlich. Dem hohen GC-Gehalt der tripelhelikalen Regionen steht durchwegs ein weitaus niedrigerer GC-Anteil der "wobble"-Positionen gegenüber. Für die weniger GC-reichen Genabschnitte des C-Terminus dagegen ist die gegenläufige Tendenz zu beobachten.

Amino säure	Codon	TRIPHELIX			NC1 DOMÄNE		
		al(IV) 'Mensch'	al(IV) 'Maus'	a2(IV) 'Maus'	al(IV) 'Mensch'	al(IV) 'Maus'	a2(IV) 'Maus'
Gly	GCG	0.18	0.17	0.23	0.35	0.29	0.22
Gly	GGA	0.28	0.33	0.34	0.06	0.12	0.11
Gly	GGT	0.30	0.30	0.13	0.06	0.24	0.17
Gly	GGC	0.25	0.20	0.29	0.53	0.35	0.50
Glu	GAG	1.00	0.67	0.63	0.73	0.70	0.73
Glu	GAA	0.00	0.33	0.38	0.27	0.30	0.27
Asp	GAT	0.56	0.57	0.29	0.40	0.17	0.25
Asp	GAC	0.44	0.43	0.71	0.60	0.83	0.75
Val	GTG	0.17	0.38	0.30	0.70	0.50	0.89
Val	GTA	0.00	0.13	0.20	0.00	0.10	0.00
Val	GTT	0.50	0.25	0.30	0.10	0.20	0.00
Val	GTC	0.33	0.25	0.20	0.20	0.20	0.11
Ala	GCG	0.00	0.00	0.05	0.18	0.12	0.06
Ala	GCA	0.25	0.00	0.21	0.29	0.29	0.24
Ala	GCT	0.25	1.00	0.53	0.18	0.24	0.35
Ala	GCC	0.50	0.00	0.21	0.35	0.35	0.35
Arg	AGG	0.00	0.00	0.18	0.23	0.23	0.00
Arg	AGA	1.00	1.00	0.18	0.31	0.38	0.00
Ser	AGT	0.00	0.00	0.05	0.12	0.08	0.09
Ser	AGC	0.00	0.00	0.30	0.32	0.33	0.50
Lys	AAG	0.18	0.44	0.48	0.80	0.80	0.88
Lys	AAA	0.82	0.56	0.52	0.20	0.20	0.13
Asn	AAT	1.00	0.00	0.00	0.44	0.20	0.25
Asn	AAC	0.00	0.00	1.00	0.56	0.80	0.75
Met	ATG	1.00	1.00	1.00	1.00	1.00	1.00
Ile	ATA	0.13	0.00	0.17	0.27	0.10	0.10
Ile	ATT	0.50	0.33	0.26	0.36	0.40	0.10
Ile	ATC	0.38	0.67	0.57	0.36	0.50	0.80
Thr	ACG	0.00	0.00	0.18	0.27	0.27	0.46
Thr	ACA	0.50	0.75	0.29	0.20	0.27	0.08
Thr	ACT	0.25	0.00	0.24	0.00	0.00	0.08
Thr	ACC	0.25	0.25	0.29	0.53	0.47	0.38
Trp	TGG	1.00	0.00	1.00	1.00	1.00	1.00
End	TGA	0.00	0.00	0.00	0.00	0.00	1.00
Cys	TGT	0.00	0.00	0.00	0.42	0.33	0.17
Cys	TGC	0.00	0.00	0.00	0.58	0.67	0.83
End	TAG	0.00	0.00	0.00	0.00	0.00	0.00
End	TAA	0.00	0.00	0.00	1.00	1.00	0.00
Tyr	TAT	0.00	0.00	0.50	0.00	0.22	0.27
Tyr	TAC	1.00	1.00	0.50	1.00	0.78	0.73
Leu	TTG	0.20	0.11	0.12	0.21	0.13	0.05
Leu	TTA	0.00	0.00	0.08	0.00	0.00	0.00
Phe	TTT	0.75	0.60	0.36	0.45	0.27	0.22
Phe	TTC	0.25	0.40	0.64	0.55	0.73	0.78
Ser	TGG	0.00	0.00	0.00	0.08	0.00	0.05
Ser	TCA	0.00	0.50	0.10	0.08	0.04	0.09
Ser	TCT	0.67	0.25	0.45	0.20	0.17	0.09
Ser	TCC	0.33	0.25	0.10	0.20	0.38	0.18
Arg	CGG	0.00	0.00	0.47	0.08	0.08	0.13
Arg	CGA	0.00	0.00	0.00	0.08	0.00	0.00
Arg	CGT	0.00	0.00	0.18	0.08	0.15	0.25
Arg	CGC	0.00	0.00	0.00	0.23	0.15	0.63
Gln	CAG	0.56	0.55	0.68	0.50	0.63	0.78
Gln	CAA	0.44	0.45	0.32	0.50	0.38	0.22
His	CAT	0.00	0.00	0.33	0.25	0.25	0.00
His	CAC	0.00	0.00	0.67	0.75	0.75	1.00
Leu	CTG	0.20	0.33	0.35	0.50	0.47	0.50
Leu	CTA	0.00	0.00	0.08	0.00	0.00	0.15
Leu	CTT	0.30	0.22	0.23	0.14	0.13	0.00
Leu	CTC	0.30	0.33	0.15	0.14	0.27	0.30
Pro	CCG	0.09	0.10	0.05	0.13	0.19	0.33
Pro	CCA	0.40	0.37	0.47	0.25	0.25	0.20
Pro	CCT	0.36	0.32	0.33	0.25	0.06	0.20
Pro	CCC	0.16	0.22	0.16	0.38	0.50	0.27

Tab. 10: Codongebrauch bei al(IV) von Mensch und Maus und a2(IV) von Maus. Die Werte sind für die NC1-Region und den tripelhelikalen Bereich getrennt aufgelistet.

Im tripelhelikalen Bereich wurden bei al(IV) von Maus 165 Codons analysiert, bei al(IV) von Mensch (Pihlajaniemi et al., 1985)

TEIL III: Aufklärung der 3'-nichttranslatierten Sequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$ und Charakterisierung der mRNAs

1. Kurze Einleitung

Über die Funktionen der 3'-nichtcodierenden Regionen von mRNAs bzw. entsprechender Genabschnitte ist bisher erst wenig bekannt. Es häufen sich jedoch in letzter Zeit die Hinweise darauf, daß auch diese Regionen eine Rolle bei der Regulation der Genexpression spielen können (siehe Diskussion). Daher war die Aufklärung der Primärstrukturen der 3'-nichttranslatierten Regionen von $\alpha 1(IV)$ und $\alpha 2(IV)$ von Interesse.

Zudem stellen die nichttranslatierten Regionen einen weiteren klar umrissenen Abschnitt der Typ IV-Gene bzw. der mRNA dar. Somit sollten Informationen darüber gewonnen werden, in welchem Ausmaß in diesem 3'-terminalen Genabschnitt - im Vergleich zu den codierenden Sequenzen - Homologien zwischen $\alpha 1(IV)$ und $\alpha 2(IV)$ ausgeprägt sind.

Bei den interstitiellen Kollagenen hatte sich gezeigt, daß hier die 3'-nichttranslatierten Regionen strukturelle Besonderheiten aufweisen. Sie sind im Vergleich zu anderen Genen ungewöhnlich lang (Yaffe et al., 1985), und zwar bis zu ca. 1.3 kb. Außerdem ist bei den interstitiellen Kollagenen, soweit untersucht, eine Längenheterogenität der mRNA festzustellen, und zwar sind jeweils zwei oder auch drei Transkripte nachweisbar. In allen Fällen konnte diese Längenheterogenität der mRNA mit dem Auftreten multipler Polyadenylierungssignale (AAUAAA) in der 3'-nichtcodierenden Region korreliert werden (z.B. Aho et al., 1983; Merlino et al., 1983).

Somit war von Interesse, ob auch Typ IV Kollagen derartige Besonderheiten aufweist. Erste Northernhybridisierungen gaben Hinweise darauf, daß sich die 3'-nichttranslatierten Bereiche von $\alpha 1(IV)$ und $\alpha 2(IV)$ ebenfalls über 1.3 bis 1.5 kb erstrecken dürften. Zudem zeigte sich, daß auch bei $\alpha 1(IV)$ mindestens zwei mRNA-Banden vorliegen, und zwar im Bereich von ca. 5.4 und 6.7 kb (siehe Teil I, 2., S. 28). Bei $\alpha 2(IV)$ dagegen konnte nur eine mRNA-Spezies beobachtet werden.

Es ist schon lange bekannt, daß die Konsensussequenz AAUAAA Signalcharakter für die Auslösung von Prozessen wie "splicing" am 3'-Ende und anschließende Polyadenylierung eukaryotischer mRNA hat (Proudfoot und Brownlee, 1976; Fitzgerald und Shenk, 1981). Bei einigen Genen ist kurz vor der Poly(A)-Stelle statt der Konsensussequenz AAUAAA eine modifizierte Form, wie z.B. CAUAAA (Birnstiel et al., 1985) präsent. Man geht davon aus, daß diese modifizierten Hexamere in der jeweiligen mRNA ebenfalls Signalcharakter besitzen; über die Effizienz dieser Wirkung ist jedoch wenig bekannt.

Die Sequenz AAUAAA tritt auch in codierenden Bereichen auf oder in noch nicht herausgeschnittenen Introns der Primärtranskripte, ohne jedoch dort eine Signalfunktion auszuüben. Dies deutet daraufhin, daß die Präsenz des Hexamers AAUAAA zwar notwendig aber nicht ausreichend ist (McDevitt et al., 1984). Folglich müssen weitere Sequenzen bzw. Faktoren an der Auslösung des 3'-"processing" beteiligt sein. Aus der Literatur sind seit kurzem zwei Sequenzabschnitte bekannt, die gleichfalls in die Regulation des 3'-"processing" der mRNA involviert sein sollen. Es handelt sich hier um die Sequenz CAU^U_CUG, die bei vielen mRNAs kurz vor oder kurz hinter der Polyadenylierungsstelle gelegen ist (Berget, 1984). Diese Sequenz ist, wie auch das Hexamer AAUAAA, komplementär zu

einem Bereich der RNA der U4 snRNPs, von denen angenommen wird, daß sie am 3'-"processing" der mRNA beteiligt sind (Berget, 1984). Desweiteren wurden kurze G/U-reiche Sequenzabschnitte identifiziert, die bei vielen mRNAs ca. 30 bp 3' von der AAUAAA-Sequenz entfernt liegen und möglicherweise durch Ausbildung von Sekundärstrukturen als Signal wirken (Birnstiel et al., 1985).

Die Präsenz der genannten Konsensussequenzen in der Umgebung von AAUAAA-Hexameren kann also darauf hindeuten, daß es sich um funktionelle Polyadenylierungsstellen handelt. Durch die mehr oder weniger starke Konservierung der Konsensussequenzen könnten verschiedene Signalstärken erreicht werden.

In keinem der aus der EHS-cDNA-Genbank isolierten Klone (p5, p9, pAIIa, pAIIb; Teil I) war die gesamte 3'-nichttranslatierte Region enthalten, obwohl die cDNA-Synthese am Poly(A)-Schwanz der mRNA mit Oligo(dT) als Primer gestartet worden war. Die vermuteten Ursachen hierfür (unvollständige Synthese des zweiten Stranges und anschließender S1-Nuklease-Verdau) wurden bereits in Teil I, 3.2. (S. 37) ausgeführt. Zur Vervollständigung der 3'-nichtcodierenden Bereiche von $\alpha 1(IV)$ und $\alpha 2(IV)$ wurden daher auf andere Weise konstruierte cDNA-Genbanken eingesetzt, und zwar war hier die Synthese analog zur Okayama-Berg-Methode (1982) erfolgt (Oberbäumer, 1986). Bei dieser Methode ist der Primer (dT)_n an den Vektor gebunden, die cDNA-Synthese findet also direkt am Vektor statt. Daher sollten alle integrierten cDNAs das 3'-Ende mit Anteilen des Poly(A)-Schwanzes enthalten. Zudem bietet diese Methode den Vorteil, daß die cDNA gerichtet im Plasmid inseriert ist.

Als Vektoren wurden die Plasmide pUC830 und pUC931 (Oberbäumer, 1986) eingesetzt, bei denen es sich um pUC8- und pUC9-Derivate mit den Polylinkerregionen (Tab.4, S. 33) der M13-Vektoren tg130 (pUC830) und tg131 (pUC931) handelt. Die

cDNA wurde jeweils zwischen die Kpn I- und die Eco RV-Stelle der Polylinker inseriert, wobei (dT)_n beim Vektor pUC830 an die Kpn I-Stelle angehängt wurde, beim Vektor pUC931 an die Eco RV-Stelle. Die Konstruktionsweise der cDNA ist in Abbildung 21 schematisch dargestellt. Für die Transformation war der E. coli-Stamm DH1 (Hanahan, 1983) eingesetzt worden.

Als Matrize für die cDNA-Synthese wurde RNA aus PYS-2-Zellen eingesetzt. Nachdem für $\alpha 1(IV)$ zwei verschieden lange mRNAs (5.4 kb und 6.7 kb) beobachtet worden waren (Teil I, 2., S. 28), sollten diese beiden Transkripte getrennt untersucht werden. Um hier gezielt die entsprechenden cDNA-Varianten zu erhalten, wurde Poly(A)+RNA durch einen Sucrose-Gradienten nach Größe aufgetrennt, und verschiedene Fraktionen als Matrize für die cDNA-Synthese eingesetzt. Es wurden zwei Genbanken cI und cII (Vektor: pUC931) mit folgenden RNA-Fractionen hergestellt:

cI: RNA aus dem Bereich zwischen der 18S und der 28S rRNA

Diese Fraktionen enthalten bevorzugt die kleine Variante (5.4 kb) der $\alpha 1(IV)$ -mRNA (Laurent et al., eingereicht) so daß hier $\alpha 1$ -Klone mit einer kurzen 3'-nichttranslatierten Region zu erwarten waren.

cII: RNA größer als die 28S rRNA

Da in diesem Bereich die 6.7 kb-mRNA von $\alpha 1(IV)$, sowie die $\alpha 2$ -mRNA von 6.5 kb auftreten, sollten entsprechenden cDNA-Klone die noch unbekannten Sequenzen des 3'-Bereichs der nichttranslatierten Regionen von $\alpha 1(IV)$ und $\alpha 2(IV)$ enthalten. Aus dieser Genbank cII war schon der $\alpha 2$ -Klon pAIIc (Abb. 13, S. 47) isoliert worden.

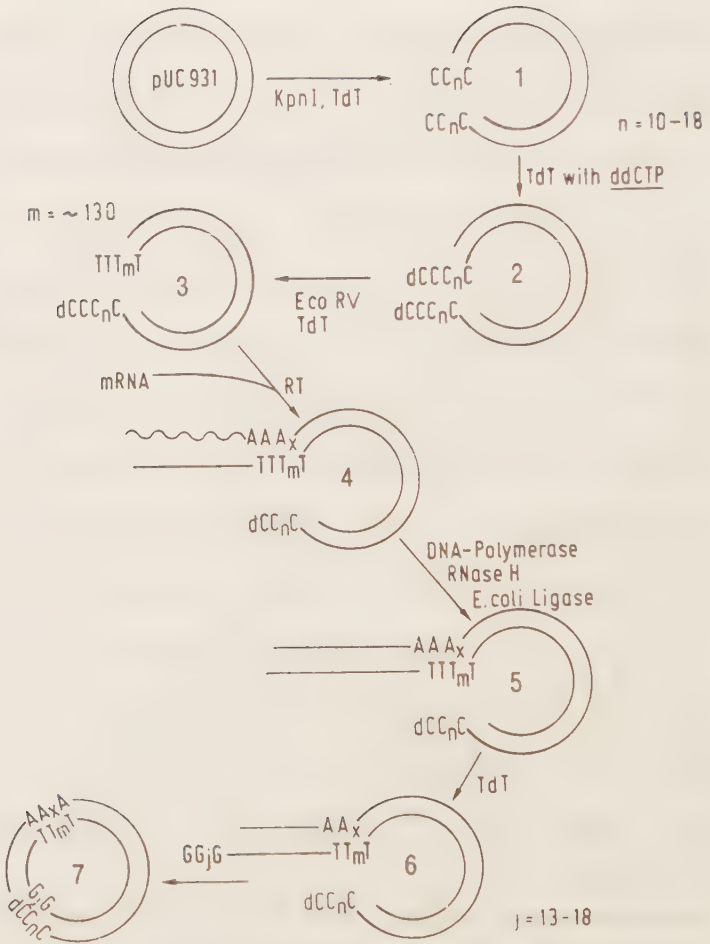


Abb. 21: Konstruktion der cDNA-Genbanken cI und cII (Oberbäumer, 1986)

- 1) Linearisierung des Vektors mit Kpn I, "tailing" mit dCTP durch terminale Transferase.
- 2) Blockierung der Enden für weitere Transferasereaktionen durch Addition von ddCTP
- 3) Entfernung des dC-Schwanzes auf einer Vektorseite durch Eco RV, Addition von dTTP durch terminale Transferase
- 4) Zugabe der RNA, Bindung des Poly(A)-Schwanzes der RNA an den dT-"tail" des Vektors, Synthese des ersten Stranges durch Reverse Transkriptase
- 5) "nick"-Bildung in der RNA durch RNase H, Synthese des zweiten Stranges durch DNA-Polymerase an RNA-Fragmenten als Primer, Verknüpfung der cDNA-Fragmente durch E. coli Ligase
- 6) Addition von dGTP an den komplementären Strang
- 7) Ringschluß durch Hybridisierung von dC- und dG-"tails"

2. Vervollständigung der 3'-nichttranslatierten Regionen von $\alpha 1(IV)$ und $\alpha 2(IV)$

2.1. Isolierung und Charakterisierung der $\alpha 1(IV)$ -spezifischen cDNA-Klone pAI4, pAI5 und pAI6

Die cDNA-Genbanken cI und cII wurden mit dem $\alpha 1(IV)$ -spezifischen Oligonukleotid Oligo I (Tab.3, S. 19) hybridisiert, das eine eindeutige Sequenz aus der NC1-Domäne (Aminosäurepos. 36-42 in Abb.14, S. 49) enthielt. Dies führte zur Isolierung des Klons pAI4 aus der Genbank cI und des Klons pAI5 aus der Genbank cII (Abb.22). Das Insert von pAI4 hatte eine Länge von ca. 1.5 kb, das von pAI5 ca. 2.4 kb. Durch Restriktionskartierung konnte die Lage beider Klone in Bezug zu dem bereits sequenzierten $\alpha 1(IV)$ -Klon pFAC festgelegt werden (Abb.22). Klon pAI5 trägt gegenüber pFAC, der ca. 750 bp der nichttranslatierten Region enthielt, ca. 500 bp weiterer bisher unbekannter Sequenz bei. pAI4 endet kurz nach einer Sau 3AI-Stelle, die auch in Klon pFAC enthalten ist.

pAI4 stammte aus der Genbank cI, für deren Synthese die RNA-Fractionen eingesetzt worden waren, die bevorzugt die 5.4 kb-Variante der $\alpha 1(IV)$ -mRNA enthalten. Die Lokalisation des 3'-Endes von pAI4 (Abb.22) war ein deutlicher Hinweis darauf, daß die cDNA dieses Klons an einer 5.4 kb-mRNA extendiert worden war. 3' von der erwähnten Sau 3AI-Stelle, in deren Bereich pAI4 endet, ist nämlich eine AATAAA-Konsensussequenz lokalisiert (Pos. 267 in Abb.25, S. 96), die als auslösendes Signal für die Entstehung der 5.4 kb-RNA angesehen wird (siehe auch Teil III,4., S. 104). Demnach müßte Klon pAI4 wenige Basen 3' von der fraglichen Polyadenylierungsstelle enden. Um dies durch Sequenzierung nachzuweisen, wurde versucht, das betreffende terminale Fragment 3' von der Sau 3AI-Stelle in M13-Vektoren subzuklonieren, was jedoch leider mißlang. Die Ursache hierfür war möglicherweise, daß dieses terminale Fragment fast ausschließlich aus Poly(A)-Sequenzen (ca. 130 bp) besteht und deshalb vom M13-Phagen nicht akzeptiert wurde.

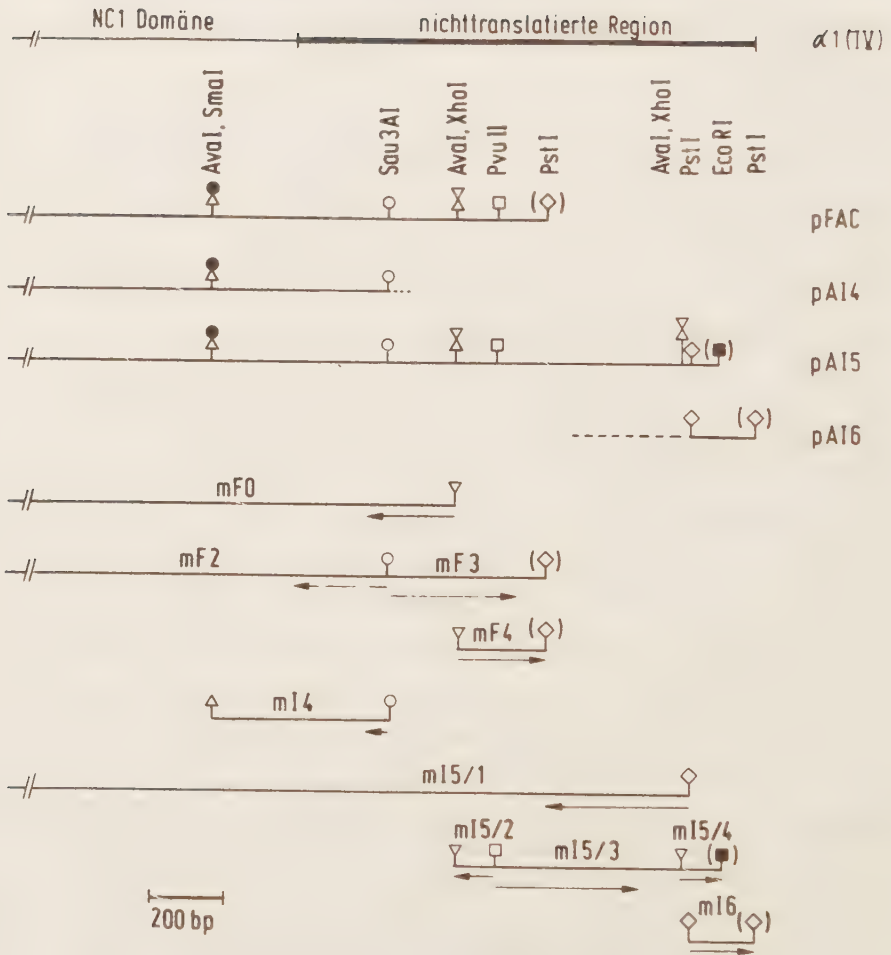


Abb. 22: Restriktionskarte der Klone pA14, pA15, pA16 und pFAC und Sequenzierungsstrategie. mF0 bis mF4 sind M13-Subklone (M13tg130 und tg131) von Klon pFAC, mI5/1 bis mI5/4 entsprechend von pA15, sowie mI4 und mI6 von pA14 bzw. pA16. Die Klammern bezeichnen Restriktionsstellen, die aus dem jeweiligen Vektor stammen.

Zur Sequenzierung des in Klon pA15 enthaltenen Abschnitts unbekannter nicht-translatierter Sequenz wurden die M13-Subklone m15/1 bis m15/4 (Abb.22) konstruiert. Das m15/3 entsprechende Pvu II/Xho I-Fragment war mit 500 bp zu lang, um beide Stränge durchgehend sequenzieren zu können. Leider ist in m15/3 auch keine geeignete Restriktionsstelle vorhanden, durch deren Spaltung kürzere Teilfragmente subkloniert und dann sequenziert werden könnten.

Subklon m15/4 enthielt mit ca. 80 bp das 3'-Ende von pA15. Die Sequenzierung des komplementären Stranges vom 3'-Ende aus zur Absicherung der Sequenzdaten war hier wegen des dort lokalisierten langen Poly(A)-Schwanzes nicht möglich. Somit wurde vom 80 bp langen 3'-Ende von pA15 nur ein Strang sequenziert.

Die Restriktionskartierung und Sequenzierungsstrategie (ausschließlich Sequenzierung von M13-Subklonen nach Sanger) der Klone pA14 und pA15 ist in Abb.22 dargestellt, zusammen mit dem überlappenden Bereich von Klon pFAC.

Die Sequenzdaten aus der Analyse der Klone pFAC und pA15 ergaben zusammen eine Länge von 1177 bp für die nichttranslatierte Region von $\alpha 1(IV)$.

Auch hier zeigte sich das gleiche Phänomen wie schon bei dem $\alpha 2$ -Klon pA11c (Abb.13, S. 47), daß nämlich das vermeintliche 3'-Ende der nichttranslatierten Sequenz nicht durch ein kurz davor liegendes Polyadenylierungssignal (AATAAA) gekennzeichnet ist. Da sowohl pA15 als auch pA11c Anteile des Poly(A)-Schwanzes enthalten, war man davon ausgegangen, daß in diesen Klonen jeweils das 3'-Ende der nichttranslatierten Sequenz vollständig enthalten sein muß. Umso überraschender war das Fehlen von 3'-terminalen Polyadenylierungssignalen in beiden Klonen, deren Präsenz doch Voraussetzung für das Anheften des Poly(A)-Schwanzes an die mRNA sein sollte. Diese Befunde waren zunächst unerklärlich. Einen Hinweis darauf, daß zumindest im $\alpha 1$ -Klon pA15 doch nicht das 3'-Ende der nicht-translatierten Region enthalten ist, brachte schließlich der Vergleich mit entsprechenden Humansequenzen, die inzwischen veröffentlicht worden waren

(Pihlajaniemi et al., 1985). Hier zeigte sich, daß zwischen beiden 3'-Abschnitten hochgradige Sequenzhomologien bestehen (siehe auch Abb.26, S. 98). Auffällig war jedoch, daß am 3'-Ende der Humansequenz zusätzlich ca. 100 bp mit einer AATAAA-Sequenz kurz vor dem Poly(A)-Schwanz vorhanden sind, und daß ein dazu homologer Abschnitt in Klon pA15 fehlte. Da bei der Northernhybridisierung (Abb.28, S. 106) keine deutliche Längendifferenz zwischen den längsten Varianten der $\alpha 1(IV)$ -mRNA von Mensch und Maus zu beobachten war, wurde aus diesen Befunden abgeleitet, daß auch die nichttranslatierte Sequenz von $\alpha 1(IV)$ der Maus eine Länge von ca. 1300 bp haben muß, diese jedoch trotz Anteilen des Poly(A)-Schwanzes nicht vollständig in Klon pA15 enthalten ist.

Die vermutete Ursache für dieses Phänomen wird im Folgenden dargelegt:

Bei der Konstruktion der cDNA-Genbank cII, aus der die Klone pA15 ($\alpha 1$) und pA11c ($\alpha 2$) stammen, war der linearisierte Vektor auf der Eco RV-Seite mit dT-Schwänzen von ca. 130 Basen versehen worden (Schritt 3 in Abb.21). Zudem war die Hybridisierung der dT-Sequenz an den Poly(A)-"tail" der mRNA unter nicht stringenten Bedingungen (30 min bei 0 °C) erfolgt. Der Publikation von Mercer und Wake (1985) ist zu entnehmen, daß die Poly(A)-Schwänze neusynthetisierter eukaryotischer mRNA rasch von 200 bis 300 Basen auf 50 bis 60 Basen reduziert werden. Wenn dies auch für die Typ IV mRNAs zutrifft, dann waren bei der cDNA-Synthese viel längere dT-Schwänze vorhanden, als für die Bindung der Poly(A)-Schwänze der mRNA benötigt wurden (bei Schritt 4 in Abb.21). Es wäre daher denkbar, daß ungepaarte überstehende dT-Schwänze teilweise mit der anschließenden A-reichen nichttranslatierten Region der mRNA hybridisiert haben, und daß daher die Synthese der cDNA durch Reverse Transkriptase (Schritt 4, Abb.21) erst in einem gewissen Abstand vom 3'-Ende der RNA begonnen hat. Dies hatte den Verlust der 3'-Abschnitte der Sequenzen zur Folge gehabt, da bei der Synthese des zweiten Stranges durch DNA-Polymerase (Schritt 5) in der betreffenden Region (dA)_n an (dT)_n als Matrize eingebaut worden wäre.

Diese Erklärung für das Fehlen des tatsächlichen 3'-Endes kann gleichermaßen für den $\alpha 1$ -Klon pAI5, eventuell für Klon pAI4, wie auch für den $\alpha 2$ -Klon pAIIC gelten, da diese Klone aus den Genbanken cI und cII stammen, deren cDNA unter den gleichen Bedingungen synthetisiert worden war.

Nachdem also mehr oder weniger erwiesen war, daß weder pAI5 ($\alpha 1$) noch pAIIC ($\alpha 2$) den 3'-Endbereich der jeweiligen mRNA enthielten, wurde eine weitere PYS-2-Genbank konstruiert (PG3, Oberbäumer, 1986). Bei dieser Genbank PG3 hatten die dT-Schwänze eine Durchschnittslänge von nur 40 bis 80 bp. Die Bindung des dT-Stranges an den Poly(A)-Schwanz der mRNA wurde außerdem unter stringenteren Bedingungen als bei cI und cII, nämlich 10 min bei 56 °C durchgeführt. Aufgrund dieser beiden Modifikationen sollte hier der oben genannte Effekt der Heteroduplexbildung nicht auftreten. Ansonsten war PG3 nach dem gleichen Prinzip wie die beiden anderen Genbanken konstruiert worden (Abb.21), und zwar mit dem Vektor pUC830 und mit RNA größer als die 28S rRNA.

Um aus dieser cDNA-Genbank PG3 Klone mit den 3'-terminalen Sequenzen von $\alpha 1$ (IV) und $\alpha 2$ (IV) zu isolieren, wurde gleichzeitig mit $\alpha 1$ (IV)- und $\alpha 2$ (IV)-Fragmenten, also einem Probengemisch, hybridisiert. Für $\alpha 1$ (IV) wurde das Fragment mI5/3 (Abb.22) nicktranslatiert und eingesetzt, für $\alpha 2$ (IV) das Fragment mc4 (Abb.24). Die siebzehn positiven Klone ($\alpha 1$ und $\alpha 2$) aus dieser Doppelhybridisierung wurden auf einen Nitrozellulosefilter gepickt und angezogen. Jeweils ein Replika dieses Filters wurde nach Lysis der Bakterien (siehe Methodenteil, S. 147) abermals hybridisiert, diesmal jedoch getrennt für $\alpha 1$ (IV) und $\alpha 2$ (IV). Wie die entsprechenden Autoradiogramme in Abb.23 zeigen, konnten für $\alpha 1$ (IV) und $\alpha 2$ (IV) positive Klone isoliert werden. Die Signale sind klar unterscheidbar - es ist also keine Kreuzhybridisierung festzustellen, was auch aufgrund der differierenden Sequenzen von $\alpha 1$ (IV) und $\alpha 2$ (IV) in diesem Bereich (siehe Teil III,3., S. 95) ausgeschlossen werden konnte.

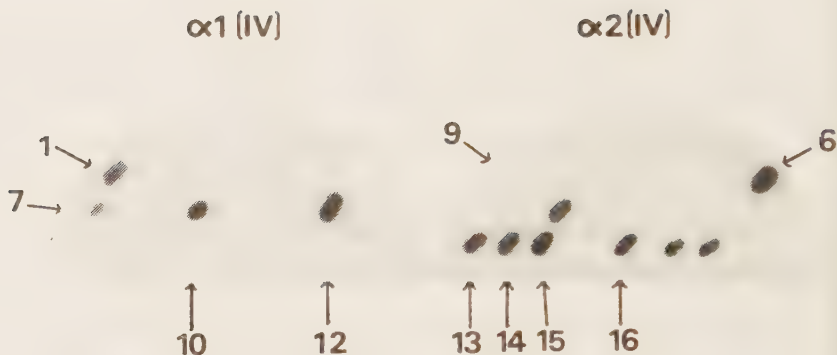


Abb.23: Differenzierung zwischen $\alpha 1$ - und $\alpha 2$ -spezifischen cDNA-Klonen, die gemeinsam aus der PYS-2-Genbank PG3 isoliert worden waren. Ein Duplikat des Nitrozellulosefilters mit den gepickten Klonen wurde mit dem nicktranslatierten Fragment m15/3 ($\alpha 1$) hybridisiert, das andere mit dem nicktranslatierten Fragment mc4 ($\alpha 2$).

Die vier $\alpha 1$ -Klone, die in Abb.23 die stärksten Signale zeigten (1, 7, 10 und 12) und daher keine Mischkolonien darstellten sollten, wurden angezüchtet. Die isolierten Plasmide wurden mit dem Restriktionsenzym Pst I geschnitten, um festzustellen, welcher dieser vier Klone das langste 3'-terminale Pst I-Fragment (Abb.22) enthält. Die Fragmentgrößen lagen zwischen 250 bp und 280 bp, was durch unterschiedliche Anteile des Poly(A)- bzw. dT-Schwanzes verursacht werden kann. Aus dem Klon mit dem langsten 3'-terminalen Pst I-Fragment (280 bp), genannt pA16 (Abb.22), wurde das betreffende Fragment isoliert, in M13tg130 subkloniert, und der so erhaltene Subklon m16 (Abb.22) sequenziert.

pA16 enthält den gesuchten Endabschnitt der nichttranslatierten Region von $\alpha 1(IV)$, der 101 bp lang ist. Erkennbar ist dies daran, daß 18 bp vor dem Ende das postulierte AATAAA-Signal auftritt, und daß der 101 bp lange Endabschnitt

einschließlich dieser Konsensussequenz fast identisch ist mit der bereits bekannten Humansequenz (Abb.26, S. 98).

Überraschend war, daß der Beginn (Pos.1178 in Abb.25, S. 96) dieses 101 bp langen Endabschnitts durch eine Folge von 15 dA-Resten gebildet wird. Ein dazu homologer Abschnitt fehlt in der α 1-Humansequenz (Abb.26, S. 98). Klon pA15 endet exakt 5' vor dieser Oligo(dA)-Sequenz. Dieser Befund unterstützt die bereits oben genannte Hypothese, daß das 3'-Ende der nichttranslatierten Sequenz in Klon pA15 fehlt, weil die zu langen dT-Schwänze mit diesem Endabschnitt einen Heteroduplexbereich ausgebildet haben. Der 15 Basen lange Oligo(A)-Abschnitt in der mRNA hat diesen Vorgang offensichtlich begünstigt, bzw. erst ausgelöst.

2.2. Identifizierung des 3'-Endes der nichttranslatierten Region von α 2(IV)

Wie im vorherigen Abschnitt beschrieben, wurde die PYS-2-Genbank PG3 gleichzeitig mit einer α 1- und einer α 2-spezifischen Probe hybridisiert, um für α 1(IV) und α 2(IV) Klone mit dem 3'-Terminus der nichttranslatierten Region zu identifizieren. Bei den sechs α 2(IV)-Klonen mit den intensivsten Signalen (6, 9, 13, 14, 15 und 16 in Abb.23) wurden wiederum die Größe des jeweiligen 3'-terminalen Pst I-Fragments (Abb.24) bestimmt. Hier gab es eine Streuung zwischen 390 bp und 450 bp. Der Klon mit dem längsten 3'-Bereich wurde selektioniert (pAIIId, Abb.24). Das entsprechende Pst I-Fragment von pAIIId wurde isoliert, mit Pvu II nachgeschnitten und das resultierende Pvu II/Pst I-Fragment in M13tg131 subkloniert. Aus unerfindlichen Gründen ergaben die mehrmaligen Versuche, diesen M13-Subklon zu sequenzieren, keine lesbaren Sequenzen. Daher wurde dieses Pvu II/Pst I-Fragment in das Plasmid pUC9 subkloniert (pd1, Abb.24) und durch Doppelstrangsequenzierung nach der Didesoxymethode (Chen und Seeburg, 1985) erfolgreich analysiert. Somit lagen schließlich auch für das 3'-Ende von α 2(IV) die Sequenzdaten vor. Der zuvor fehlende Abschnitt, lediglich 20 bp, ergänzt die nichttrans-

latierte Sequenz von $\alpha 2(IV)$ auf insgesamt 969 bp. Das postulierte Polyadenylierungssignal AATAAA beginnt in Position 951 (Abb.27, S. 102).

Klon pAIIc endete direkt vor dieser Sequenz. Offensichtlich hatte auch hier der überlange dT-Schwanz des Vektors mit der A-reichen Region des Polyadenylierungssignals am 3'-Terminus der mRNA hybridisiert und die Synthese des 3'-Endbereichs verhindert.

Mit einer Länge von 969 bp ist die nichtcodierende Region von $\alpha 2(IV)$ um 330 bp kürzer als bei $\alpha 1(IV)$. Dieses Resultat war überraschend, da bei Northernhybridisierungen die $\alpha 2$ -mRNA und die längste Variante der $\alpha 1$ -mRNA komigrieren (siehe Abb.28, S. 106). Die Vermutung, daß der Klon pAII d immer noch nicht das 3'-Ende enthielt, ist durch die Präsenz des Polyadenylierungssignals unwahrscheinlich.

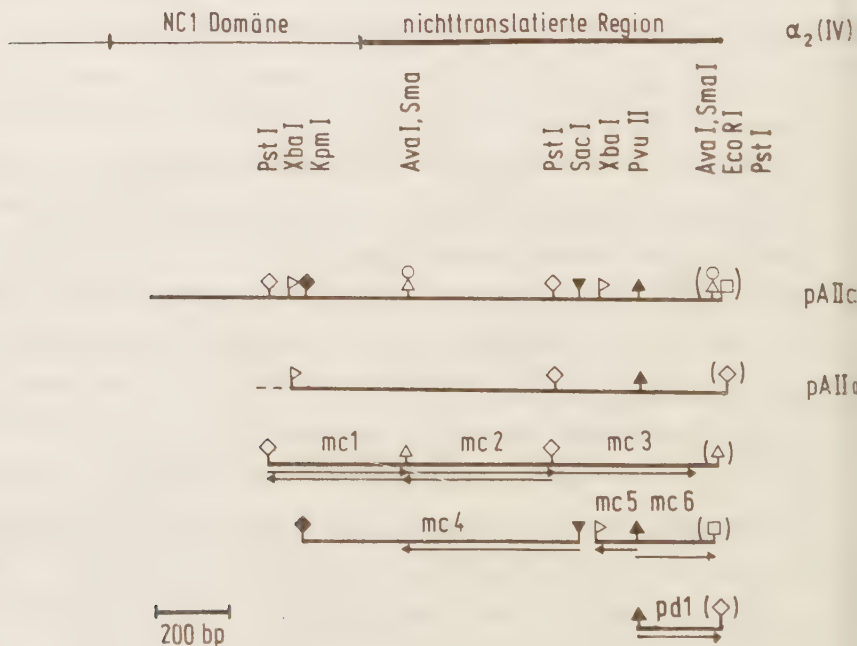


Abb 24: Lage von Klon pAII d im Vergleich zu Klon pAII c
Bei mc1 bis mc6 handelt es sich um M13-Subklone von pAII c in den M13-Vektoren tg130 und tg131. pd1 ist ein Subklon von pAII d in dem Vektor pUC9. Richtung und Länge der Sequenzierungen sind durch Pfeile angezeigt. Die eingeklammerten Restriktionsstellen werden vom jeweiligen Vektor beitrugen.

3. Die Sequenzen der 3'-nichttranslatierten Regionen von $\alpha 1(IV)$ und $\alpha 2(IV)$

Durch Sequenzierung der Klone pFAC, pAI5 und pAI6 konnte die gesamte 3'-nicht-translatierte Region von $\alpha 1(IV)$ mit einer Länge von 1278 bp aufgeklärt werden (Abb.25).

Der Bereich enthält sieben kanonische Polyadenylierungssignale (AATAAA), die in Abb.25 eingerahmt sind. 267 bp nach dem Ende der codierenden Region findet sich bereits das erste Polyadenylierungssignal. Im Bereich zwischen Position 593 und 727 treten dann gehäuft AATAAA-Sequenzen auf, insgesamt fünf an der Zahl, wobei zwei dieser Hexamere überlappen (Pos.620 bis 631). Am 3'-Ende der Sequenz (Pos.1262) schließlich erscheint die letzte AATAAA-Sequenz.

Neben der Sequenz AATAAA finden sich drei abgewandelte Formen (siehe Teil III,1., S. 83), CATAAA, AGTAAA und AATATA, die in Abb.25 durch eine gepunktete Linie gekennzeichnet sind. Die Ergebnisse von Northernhybridisierungen geben allerdings keinen Hinweis darauf, daß eines dieser Hexamere eine Signalfunktion ausübt.

Hinsichtlich der anderen potentiellen Signalsequenzen (siehe Teil III,1., S. 83) ist nach dem ersten Poly(A)-Signal in Pos. 267 die Sequenz CACTG zu finden (Pos.300), sowie kurz danach in Position 306 ein G/T-reicher Abschnitt (TGCTCTGGGG). Beide sind in Abb.25 durch eine Linie markiert. Die Präsenz dieser potentiellen Signalsequenzen ist ein weiterer Hinweis darauf, daß ein Teil der pre-mRNA von $\alpha 1(IV)$ nach dieser ersten AATAAA-Sequenz in Position 267 prozessiert wird (siehe Teil III,4., S. 103). Im Bereich der anderen AATAAA-Sequenzen bei $\alpha 1(IV)$ sind keine der zwei angeführten potentiellen Signalsequenzen zu finden, obwohl wahrscheinlich zumindest eines dieser Hexamere funktionell ist (siehe Teil III,4., S. 103). Für die letzte Polyadenylierungssequenz in Pos.1262 dürften mögliche Signalsequenzen 3' von der Schnitt- und

1	TA	ACTATGAT	GCTCGCCTCT	GCCACCGTCA	CAACATGGTG	CTACTCCCCC	50
51	CCTT	GTGCTCTC	TTCTCCCTCC	CATTGTATTT	GTAACAGCAA	TGAACCCTAG	100
101	AATT	TATATCC	TGTGTACCTC	ACTGTCCACT	ATGAAAACCA	TAAAGTGCCT	150
151	TATAG	GAATC	CGCGTAACTA	ACACACCCCG	CTTCGTTGGC	CTCTGTTTGC	200
201	TGAAG	GAGAA	GGACAGGCAG	CGGAGAGCTT	TCCATAATGG	CACATACCAA	250
251	ATGGCGCTCCTGATCC	AATAAA	AGATGAAT	TGCATTTTGC	AATCCAGTGC		300
301	ACTGATGTCT	GGGGCGAGAA	CTCCATAAGA	AACCCAAAGG	TGCTAGGGGC		350
351	GTGAGGCGCT	CTCCACACAG	CCCACAGCCC	GCTTCATCCT	GTCAACCCAG		400
401	ATGTGAGTGC	TGTCCAACCC	ACTGTCCAGT	GTTTCCACAT	CAGGTCCCTC		450
451	GAGGGGCTAT	GCAGAGAACA	GTCAAGGGAA	CTGGTTTCTT	CCACTTTCTA		500
501	CATGGAATC	CATCGTTTCA	GCCCCAAACA	CAAACCTTATG	CCCTGTGTGG		550
551	TACTATGCAG	CTGCTTTTGT	GGAAGTCATGGCTTTCCTGTGG	AATAAA	GA		600
601	TGGTCCAAAGGTTAGATGGAA	AATAAATAAA	TAAGTA	AATAAA	TATATAT		650
651	ATATATTAGT	AATTTATATA	GGTGCAGCA	ATTAGGCAGG	TCAAGTTCTA		700
701	GCGTAGAGCCTGTTGTTA	AATAAA	GCGCAC	TTGGTTTTCG	TCTCTGGAGA		750
751	AACTGTACTA	TCCTTTCACA	GGGTACCAA	CATCAGGCCA	CTGAATGTGA		800
801	AACACCAGCT	TTCCTCAAAC	TCACTCCTAA	ACCAGAGGAG	GTGTATAGAT		850
851	AGCCAAAGCC	AAACCCATTC	CTAAACATGG	TGTCTGATAA	TGAGCCCTGT		900
901	GCCTGGCGCT	TCTTGCTTCT	GCTCTGCGTG	GAGTATTTTC	AGGGTTTGCA		950
951	ACACTAACCA	CAGACTGAAT	GACTGACTTC	CCGTACGACA	GCCAAGGCCT		1000
1001	TTGTTAGCCT	CCTTCATCCT	GCATCCTGGC	AGGCATTTCA	CACCTGAGCA		1050
1051	CACACAGGCA	ACTTCCCAGC	TCCTCTCGAG	CTGTGACACA	GCCCCTGCAG		1100
1101	GGAGTGGACA	GGTGCAGGA	GGGCAGACTC	GGACCACTAT	GCTTGAAGTG		1150
1151	ACATTTCACT	AAAAGTACCC	ACCCTCCAAA	AAAAAAAAAA	AAGTTTTTAA		1200
1201	GACTACTAAG	GCCTTTTGTG	TAATTTCTTT	AAATGTGTAT	TTCTTAAGAA		1250
1251	TTCAAATTTGT	AATAAA	ACTATTTGTAT				1278

Abb.25: Sequenz der 3'-nichttranslatierten Region von α 1(IV)
 Die AATAAA-Konsensussequenzen sind eingerahmt, nichtkanonische Hexamere durch eine gepunktete Linie gekennzeichnet. Eine in Position 300 auftretende CACTG-Sequenz mit einer nachfolgenden TG-reichen Sequenz sind unterstrichen (näheres im Text).

Polyadenylierungsstelle liegen. Nachdem die pre-mRNA von $\alpha 1(IV)$ eindeutig bevorzugt nach dem 3'-terminalen Hexamer geschnitten und polyadenyliert wird (siehe Teil III,4.), wäre hier die Analyse von genomischer DNA im Hinblick auf die Identifizierung von weiteren Signalsequenzen interessant.

Für die nichttranslatierte Region von $\alpha 1(IV)$ liegen, wie bereits erwähnt, auch Sequenzdaten von Mensch (Pihlajaniemi et al., 1985) vor. Beim Vergleich der beiden Bereiche zeigt sich eine stark ausgeprägte Interspezieshomologie (Abb.26). Die Konservierung der nichtcodierenden Sequenzen von Mensch und Maus hatte den entscheidenden Hinweis geliefert, daß die Sequenzdaten aus der Analyse von Klon pA15 noch nicht das 3'-terminale Ende enthielten (Teil III,2.1., S. 89).

Die parallele Zuordnung der nichttranslatierten Sequenzen von Mensch und Maus (Abb.26) erfolgte durch Anwendung des Computerprogrammes "Bestfit" (Devereux et al., 1984). Die beiden Sequenzen sind über weite Strecken identisch, in vielen Fällen nur durch einzeln auftretende Basenaustausche unterbrochen. Trotz dieser Homologie sind immer wieder größere Deletionen jeweils in einer der beiden Sequenzen zu finden, die eine Ausdehnung von bis zu 31 bp (Pos. 1031 von Maus) erreichen. Bei $\alpha 1(IV)$ von Maus treten diese Deletionen häufiger auf, weshalb diese Sequenz insgesamt um 21 bp kürzer ist als die Humansequenz.

Hinsichtlich der Basenaustausche zwischen den beiden nichttranslatierten Regionen ist die Tendenz festzustellen, daß Positionen, die in der Maussequenz mit einem G oder C besetzt sind, bevorzugt gegen A und T in der Humansequenz ausgetauscht sind. Daraus resultiert ein signifikant niedriger GC-Gehalt bei $\alpha 1(IV)$ von Mensch (GC/AT-Verhältnis: 0.55) gegenüber $\alpha 1(IV)$ der Maus (GC/AT-Verhältnis: 0.84). Beide Werte liegen jedoch deutlich unter den entsprechenden Werten der codierenden Bereiche (Tab.9, S. 77). Im Vergleich zu anderen Kollagen-Genen, z.B. $\alpha 1(I)$ und $\alpha 2(I)$ von Huhn mit einem GC/AT-Verhältnis von 0.3 und 0.4 (Fuller und Boedtker, 1981), ist der GC-Gehalt der nichtcodierenden Regionen bei $\alpha 1(IV)$ und $\alpha 2(IV)$ signifikant höher.

M 1 TAACATGATGCTCGCCTCTGCCACCGTCACAACATGGTGTCTACTCCCCCCTTGTCTCTC
H 1 --- A CT A A T AT ----- T

61 TTCTCCCTCCCATTTGTATTGTAAACAGCAATGAACCCCTAGAATTATATCCTGTGTACCTC
48 ----- T GT C A

121 ACTGTCCACTATGAAAACCATAAAGTGCCTTATAGGAATCCGCGTAACTAACACACCCCG
97 A G TT T

181 CTTCTGTTGGCCTCTGTTTGTGAAGGAGAAGGACAGGCAGCGGAGAGCTTTCCATAATGG
157 A A AC -- A A ATA - A G

241 CACATACCAAATGGCGCTCCTGATCCATAAAGATGAATTGCATTTTGAATCCAGTGC
214 -- A TT GA T C - ATT C A

301 ACTGATGTCTGGGCGGAGAACTCCATAAGAAACCCAAA-GGTGCTAGGGGCGTGAGGCGC
271 G AA T C A G A GTGTG T

360 TCTCCACACAGCCACAGCCCGCTTCATCC-----TGTCACCCCGA
331 CT T T TTTG -- ATT T TGTATTATAATTAATTT C AC

402 TGTGAGTG-CTGTCCAACCCACTGTCCAG-TGTTCCACA-TCAGGTCCCTCGAGGGGCT
389 GA A A TT TT TAT T C A A T T ---T T

459 ATGCAGAGAACAGTCAAGGGAAGTGGTTTCTTCCACTTTCTACATGGAAGTCC-----
446 G A A T T CA T TA AAT G T TG ACC C A T A AGACTCA

512 -----ATCGTTTCAGCCCAACACAAA-----CTTATGCCCTGTGTGGTACTA
506 GTAGCCAT T CCA C C GT T TTTCTGTCT C TA T

556 TGCAGCTGCTTTTGTGGAAGTCATGGCTTTCCTGTGGAATAAAGATGGTCCAAAGGTTAG
566 CA A CAAT -----

616 ATGGAAATAAATAAATAAGTAAATAATATATATATATATTAGTAATTTATATAGGTGT
618 --- GTC T A

676 CAGCAATTAGGCAGGTCAAGTCTTAGCGTAGAGCCTGTTGT-AAATAAGCGCACTTGG
669 A G T TT ACTT CAC A TT A A

735 TTTTCGTCTCTGGAGAACTGTACTATCCTTT--CACAGGGTTACCAACATCAGGCCACT
729 T CT T A G G G AA T T TTA G CT AT T

793 GAATGTGAAACACCAGCTTTCCTCAAACCTCACTCCTAAACAGGAGGGTGTATAGATAG
789 T C T A --TGT T ---- AAAT GTG T

853 CCAAAGCCAAACCCA--TTCCTAAACATGGTGTC--TGATAATGAGCCCTGTGCCTGGCG
843 A GG A G T TA T G AA A A T T A

909 CTTCTGTCTTCTGCTCTGCGTGGAGTATTTTCAGGGTTTGCAACACTAACCACAGACTGA
903 A T T AT TA TA A TA T A

969 ATGAC-TGACTTCCCGTACGACAGCC---AAGGCC-----TTTGTAGCCTCCTTCATCC
963 T AAAAG A A TTA GTCAT CA TA T C T

1020 TGCATCCTGGC-----AGGCATTTACACCTGAG
1023 A TTGAAAAACAGCTCTGTTGAATCACAGTATC TAT G A

1049 CACACACAGGCAACTTCC-CAGCTCCTCTCGAGCTGTG-ACACAG-CCCCTGCAGGGAGT
1083 TT G C T GTG T T AT TT A T A CA

1106	GG-ACAGGT	CGCAGGAGGG	CAGACTCGGAC	CACTATGCTT	GAAAGTGACATT	TTCACTAAAA	
1143	C C TG AC		T	G C	A		
1165	GTACCCACCC	TCCAAAAA	AAAAAAGG	TTTTTAAG	ACTACTAAGG	CCTTTTGTGTAAT	
1203	-----	C	-----	C		A	
1225	TTCTTTAAAT	GTGTATTCT	TTAAGAATT	CAAATTGTA	AATAAA	ACTATTTGTATA	1278
1246							1299

Abb. 26: Parallele Anordnung der 3'-nichttranslatierten Regionen von $\alpha 1(IV)$ von Maus und Mensch. In der zweiten Zeile ist jeweils die Humansequenz (H) aufgeführt, wobei nur die zur Mausequenz (M) unterschiedlichen Nukleotidpositionen angegeben sind. Positionen mit Deletionen sind durch einen Strich gekennzeichnet, AATAAA-Konsensussequenzen sind eingerahmt.

Die Homologie der beiden $\alpha 1$ -nichttranslatierten Sequenzen von Mensch und Maus schließt die Positionen von vier Polyadenylierungssignalen mit ein (Pos. 267, 593, 719 und 1262 von Maus). Dem Hexamer in Pos. 638 (Maus) entspricht bei $\alpha 1(IV)$ von Mensch die Sequenz ATTAATA, das mit einem AATATA überlappt. Beiden Hexameren wird ebenfalls Signalcharakter für das 3'-"processing" mancher mRNAs zugeschrieben (Birnstiel et al., 1985). Die beiden überlappenden AATAAA-Sequenzen ab Pos. 622 sind in der Humansequenz nicht präsent, und auch sonst treten bei $\alpha 1(IV)$ von Mensch keine zusätzlichen AATAAA-Konsensussequenzen auf. Dem ersten Polyadenylierungssignal folgt in der Humansequenz, ebenso wie bei der Maus, ein CACTG (Pos. 270), sowie ein kurzer G/T-reicher Abschnitt (TGTGTG, Pos. 276). Insgesamt gesehen ist der flankierende Bereich dieser ersten AATAAA-Sequenz sehr stark bei Mensch und Maus konserviert. So spricht auch bei $\alpha 1(IV)$ von Mensch vieles dafür, daß diese erste AATAAA-Sequenz eine Signalfunktion ausübt (Pihlajaniemi et., 1985; Myers et al., 1986).

Den am stärksten konservierten Bereich stellt jedoch der 3'-terminale Abschnitt bei $\alpha 1(IV)$ von Mensch und Maus dar. Die jeweils letzten 60 bp sind vollkommen identisch. Dies ist insofern von Bedeutung, weil es sich hier um die flankierenden Sequenzen des 3'-terminalen Polyadenylierungssignals handelt. Dieses Hexamer hat für beide mRNAs die stärkste Signalwirkung unter den multiplen Polyadenylierungssignalen, wie die Überrepräsentierung der langsten mRNA-Spezies in Northernhybridisierungen anzeigt (Abb.28, S. 106).

Der 3'-nichttranslatierte Bereich von $\alpha 2(IV)$ wurde durch die Sequenzierung der cDNA-Klone pAIIc und pAIIId vollständig aufgeklärt. Er umfaßt 969 bp. Das Hexamer AATAAA tritt in Pos.951 kurz vor dem 3'-Ende auf (Abb.27). Daneben findet sich auch zweimal die Sequenz CATAAA. Die Ergebnisse der Northernhybridisierung geben keinen Hinweis darauf, daß die CATAAA-Sequenz in Pos.441 Signalcharakter hat. Bei dem zweiten CATAAA (Pos.947), ist sowieso eine Überlappung mit der bei $\alpha 2(IV)$ singulären AATAAA-Sequenz vorhanden, deren Präsenz hier der dominierende Faktor für die Polyadenylierung der mRNA sein dürfte.

Andere potentielle Signalsequenzen wie $CA_C^U TG$ und G/T-reiche Abschnitte konnten wiederum nur durch Analyse genomischer DNA identifiziert werden.

Der GC-Gehalt der nichtcodierenden Region von $\alpha 2(IV)$ liegt mit einem GC/AT-Verhältnis von 1.06 deutlich höher als bei $\alpha 1(IV)$ (Maus: 0.84; Mensch: 0.55).

Die auffälligste Diskrepanz beim Vergleich der nichttranslatierten Regionen der cDNA von $\alpha 1(IV)$ und $\alpha 2(IV)$ ist, daß bei $\alpha 2(IV)$ nur eine einzige AATAAA-Sequenz präsent ist, bei $\alpha 1(IV)$ dagegen sogar sieben dieser Hexamere. Auch die damit verbundene Längenheterogenität der $\alpha 1$ -mRNA ist erwartungsgemäß bei $\alpha 2(IV)$ nicht erkennbar (siehe Abb.28, S. 106). Außerdem unterscheiden sich die Längen der beiden Regionen mit einer Differenz von über 300 bp sehr deutlich.

Daher wurde analysiert, ob zumindest auf Sequenzebene Ähnlichkeiten der nicht-translatierten Regionen von $\alpha 1(IV)$ und $\alpha 2(IV)$ zu verzeichnen sind. Zu diesem Zweck wurde wiederum das Computerprogramm "Bestfit" (Devereux et al., 1984) angewendet. Die Analyse hatte zum Ergebnis, daß keinerlei Sequenzhomologien zwischen den beiden Bereichen bestehen. Lediglich ein 19 bp langer Abschnitt, unterbrochen durch drei Basenaustausche, ist bei $\alpha 1(IV)$ und $\alpha 2(IV)$ identisch. Dieser Abschnitt erstreckt sich bei $\alpha 1(IV)$ von Pos.578 bis 596, bei $\alpha 2(IV)$ von Pos.927 bis 945. In beiden Fällen ist die fragliche Sequenz kurz vor einem AATAAA-Hexamer lokalisiert ($\alpha 1$: Pos.593; $\alpha 2$: Pos.951). Ob dieser Tatsache eine Bedeutung beizumessen ist, bleibt offen.

Schon beim Vergleich der codierenden Sequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$ zeichnete sich deutlich die Tendenz ab, daß die Interspezieshomologie bei $\alpha 1(IV)$ von Mensch und Maus weit stärker ausgeprägt ist als der Verwandtschaftsgrad von $\alpha 1(IV)$ und $\alpha 2(IV)$ innerhalb der Spezies Maus (Teil II,3.1., S. 72). Besonders krass kommt diese Diskrepanz bei der Gegenüberstellung der 3'-nichttranslatierten Sequenzen von Typ IV Kollagen zum Ausdruck. Zwischen $\alpha 1(IV)$ und $\alpha 2(IV)$ von Maus besteht praktisch keinerlei Ähnlichkeit, während bei $\alpha 1(IV)$ von Mensch und Maus in hohem Ausmaß Übereinstimmung zu finden ist. Die Folgerungen, die aus diesen Befunden gezogen werden können, werden im Diskussionsteil behandelt.

1	TGAGGCTCCC	GCCCTGCCCA	CGGCAGCTGT	GTGGGAAAGG	CCACGTTGGT	50
51	GCTTATCCTT	CTGTATTGTT	TTCAACTTTG	TTTGAGACAA	TGTCTCACTT	100
101	TTCATCCCTG	CCTAGCCTGG	AACTATGTAA	CCCCGGGCTG	GCCTCCAAC	150
151	TGTGGCAAAC	CTCCTGCCTC	AGCTTCCCAG	ATGCTGGGAT	TATGGGCATG	200
201	CGCCACCACA	CCCAGCTTGG	TGCTTACTCT	TAACCTATTA	CCTCAGATGC	250
251	TGATCCCAAA	GTTGATCTTA	GTGTTTTCAA	ACTACCAAAG	CAATTCTGCA	300
301	CCAGCACTCA	TTCCAACCGT	CTGTGAGCAA	GTGGGCACTA	GCTCCTCCGG	350
351	GATCTACCAG	CCTTGGCTCA	CAGCCCTGCT	AACACCCACC	TTGGAGAGCC	400
401	AACACCTAGG	AGCAACCCAC	GGACCTCTCC	CCATGTAGGA	CATAAACATC	450
451	ATGCTACCAG	GTGTCCCCAA	CAATAGCTGT	CTCTGTGTCC	CTGTGCTTCC	500
501	CCTGGCAACA	GGCTGCTCAC	CTTCCTGCAG	TCTTCCTAAA	ATGAGGCAGA	550
551	GTCTAGGGGA	CAGCAAAGGA	ATCTATGGCC	AATGATAGCT	GAGCTCACTG	600
601	ACCCATTTC	ACCCAAGGAT	GCATGCAGTG	TGCTTCTGTG	TAGACCTGGT	650
651	TTCTTGGAGG	AGGAGATGAT	AGCACTTTTT	TCTATCAAAT	CCATAACTGC	700
701	CACCTCCCAA	GTCCCTACGT	GTTGGCCCCT	GCCTTTTAA	GACCCTGACG	750
751	GCAGCTGGCA	CACAGCTCAT	GGCTCTAGCC	AAACCCACAG	CAAAGGAGAC	800
801	ATCTCTCTAG	CCCTCACCAG	CATCTGCTGG	AAGGTACACT	TAGCAAATCA	850
851	TTGACAGTGG	CGTCCTACAA	CCTGTGTCAC	CCTAAACACC	TCCCTCCACA	900
901	ACCTGTTCCC	TGCGGCAAGC	TGTTCTATGG	CTTTTCTGTG	TAACACCATA	950
951	AATAAA	AGATATTCTAAGG	969			

Abb. 27: Sequenz der 3'-nichttranslatierten Region von $\alpha 2(IV)$

Die Signalsequenz AATAAA ist eingerahmt, zwei auftretende nicht-kanonische Hexamere (CATAAA) sind durch eine gepunktete Linie gekennzeichnet.

In den 3'-nichttranslatierten Regionen von $\alpha 1(IV)$ und $\alpha 2(IV)$ sind mehrere offene Leserahmen vorhanden. Bei $\alpha 1(IV)$ umfaßen die längsten offenen Leserahmen 55 Aminosäuren (Beginn: Pos.459), sowie 46 (Pos.131) und 42 (Pos.6) Aminosäuren. Bei $\alpha 2(IV)$ beginnt in Position 181 ein 85 Aminosäuren langer Abschnitt. Außerdem treten bei $\alpha 2(IV)$, neben einigen kürzeren offenen Leserahmen, einer mit 52 Aminosäuren (Pos.582), einer mit 47 (Pos.90) und einer mit 44 (Pos.451) Aminosäuren auf.

Für $\alpha 1(IV)$ ist auch hinsichtlich der offenen Leserahmen ein Vergleich mit den entsprechenden Humansequenzen (Pihlajaniemi et al., 1985) möglich. Da die nicht-translatierten Bereiche von Mensch und Maus auf DNA- bzw. RNA-Ebene sehr ähnlich sind, kommt die Interspezieshomologie auch in den abgeleiteten möglichen Proteinsequenzen zum Ausdruck. Dies schließt die Sequenzen einiger offener Leserahmen mit ein, deren abgeleitete Peptidsequenzen abschnittsweise bei Mensch und Maus identisch sind. Keine Übereinstimmung zeigt sich jedoch hinsichtlich Beginn und Ende, und damit auch auf die Länge der analogen Leserahmen. Demnach ergibt sich zumindest für $\alpha 1(IV)$ kein Hinweis darauf, daß einer dieser offenen Leserahmen in ein Peptid translatiert wird, da in diesem Fall die entsprechende Sequenz vermutlich in hohem Ausmaß bei Mensch und Maus konserviert wäre.

4. Charakterisierung der Typ IV mRNA

Northernhybridisierungen mit Fragmenten der isolierten Typ IV cDNA-Klone wurden durchgeführt, um für $\alpha 1(IV)$ die Längen der verschiedenen mRNAs zu bestimmen, und um diese mit den zahlreich auftretenden Poly(A)-Signalen zu korrelieren. Zudem sollte das Muster der Längenheterogenität in verschiedenen Zelllinien bzw. Geweben der Maus miteinander verglichen werden.

Außerdem sollte festgestellt werden, welche Größenunterschiede sich zwischen der längsten Variante der $\alpha 1$ -mRNA und der $\alpha 2$ -mRNA zeigen, nachdem die 3'-nicht-

translatierten Regionen in ihrer Länge um über 300 bp differierten.

Schließlich war von Interesse, ob Kreuzhybridisierung mit humaner Typ IV mRNA zu beobachten ist, und ob im Falle eines positiven Resultats Längenunterschiede zur entsprechenden Maus-RNA auftreten.

Bei der eingesetzten RNA handelt es sich um Präparationen aus der menschlichen Tumorzelllinie HT-1080, dem EHS-Maustumor, sowie aus den Mauszelllinien PYS-2 und F9 (nicht induziert). Herkunft und Eigenschaften dieser Tumorsysteme wurden in Teil I,1. (S. 17) kurz beschrieben.

Gleiche Mengen (ca. 10 µg) der verschiedenen Poly(A)+mRNA (Isolierung und Anreicherung, siehe Methodenteil, S. 138) wurden in einem denaturierenden Glyoxal-gel aufgetrennt und durch Elektroblothing auf Filter (Genescreen) transferiert (siehe Methodenteil, S. 144).

Zunächst wurde mit dem α 1-spezifischen M13-Klon m15/1 (Abb.22) hybridisiert. Dieser Klon enthält neben Anteilen der codierenden Region fast den gesamten nichttranslatierten Bereich von Klon pA15. Somit sollten alle in der Länge der nichttranslatierten Region variierenden α 1(IV)-mRNAs markiert werden.

Das Resultat dieser Hybridisierung (Abb.28A) zeigt, daß bei der PYS-2- und F9-RNA jeweils eine Bande zu sehen ist, bei der RNA aus dem EHS-Maustumor dagegen zwei Banden auftreten. Anhand von λ -Markern konnte für die obere Bande eine Länge von ca. 6.5 kb bestimmt werden. Für die zweite Bande bei der EHS-RNA wurde eine Länge von 5.5 kb ermittelt.

Bemerkenswert ist der Befund, daß das Längenmuster der α 1(IV)-RNAs abhängig vom Zelltyp variiert. In der Präparation aus dem EHS-Tumor zeigen sich zwei RNAs (5.5 kb und 6.5 kb), die in ungefähr äquimolaren Mengen vorhanden sind. In den beiden anderen RNA-Populationen dagegen ist neben der 6.5 kb-Bande keine weitere zu erkennen. Dies war zumindest für die PYS-2-RNA überraschend, da bei früheren Northernhybridisierungen von PYS-2-RNA mit einem Oligonukleotid (Oligo I, Teil I,2., S. 28) oder cDNA-Proben (Oberbäumer et al., 1985) neben einer stark

markierten 6.7 kb-RNA eine, wenn auch weitaus schwächere 5.4 kb-Bande nachweisbar war. Bei der vorliegenden Northernhybridisierung wurde eine andere RNA-Präparation der PYS-2 Zelllinie eingesetzt. Daher könnten die beobachteten Unterschiede in der Längenheterogenität bei PYS-2 Ausdruck verschiedener Wachstumsstadien der jeweils verwendeten Zellkulturen sein.

Die auftretenden Größendifferenzen der fraglichen Banden (6.7/6.5 kb und 5.4/5.5 kb) bei den verschiedenen Northernhybridisierungen dürften durch verschiedene Gelsysteme verursacht werden (Bei früheren Northernhybridisierungen mit $\alpha 1(IV)$ -spezifischen Proben war an Stelle von Glyoxal Methylquecksilberhydroxid zur Denaturierung der RNA eingesetzt worden). Die bei der vorliegenden Northernhybridisierung ermittelten Werte von 6.5 kb und 5.5 kb stimmen allerdings besser mit den Längen überein, die anhand der Sequenzdaten berechnet werden können (siehe S. 107).

Die Resultate dieser Hybridisierung (Abb.28A) zeigten also, daß das Muster der Längenheterogenität der $\alpha 1(IV)$ mRNA abhängig vom Zelltyp und eventuell vom Wachstumszustand variiert. Die mögliche Bedeutung dieses Befundes im Zusammenhang mit Regulationsvorgängen wird im Diskussionsteil behandelt.

Bei der in Abbildung 28A gezeigten Northernhybridisierung wurde parallel auch humane HT-1080-RNA analysiert. Wegen zu schwacher Markierung ist das entsprechende Bild nicht gezeigt. Es war also in nur sehr geringem Ausmaß Kreuzhybridisierung mit humaner $\alpha 1(IV)$ -mRNA zu beobachten, trotz der Sequenzhomologien zwischen Mensch und Maus. Die markierte Bande komigrierte jedoch mit der oberen Bande der $\alpha 1$ -RNA von Maus und hat demnach auch eine Länge von 6.5 kb. Für die HT-1080-Zelllinie ist bekannt, daß auch sie mindestens noch eine weitere mRNA-Spezies exprimiert (Pihlajaniemi et al., 1985).

Die Tatsache, daß die längsten $\alpha 1$ -mRNAs von Mensch und Maus komigrieren, steht im Einklang mit den Befunden, wonach bei $\alpha 1(IV)$ von Mensch und Maus eine starke Interspezieshomologie ausgeprägt ist, die die Länge der 3'-nichttranslatierten Regionen miteinschließt.

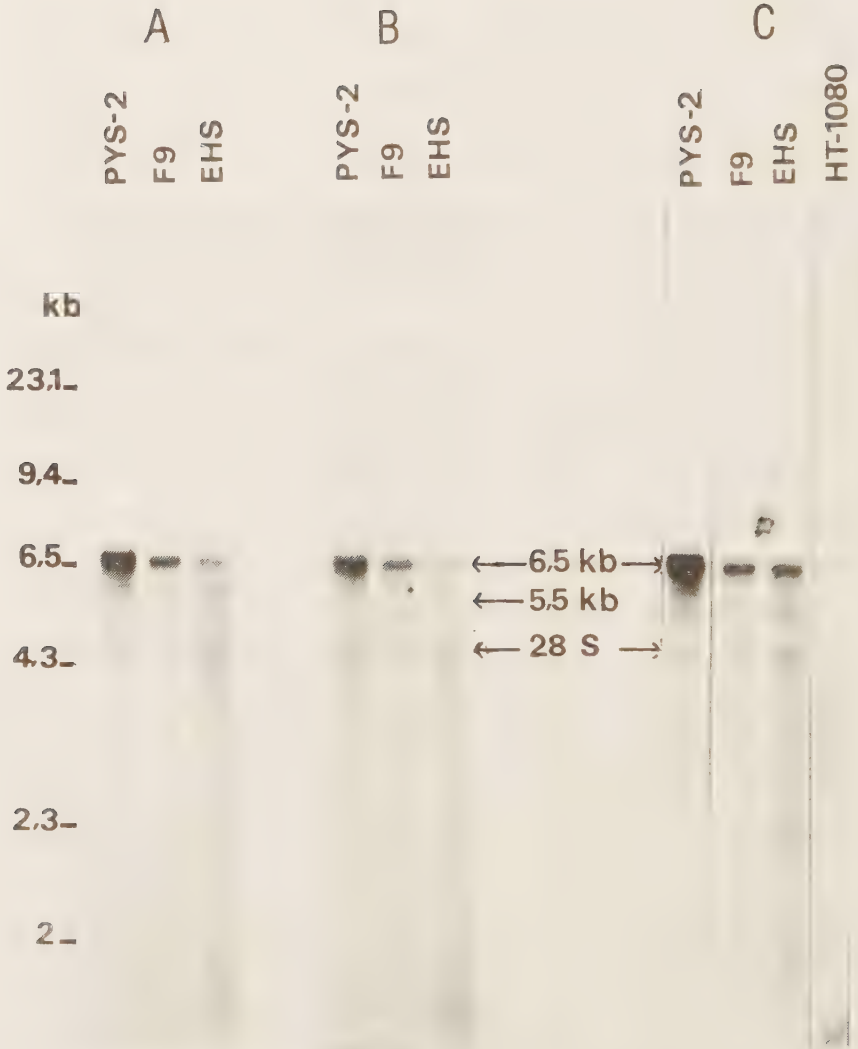


Abb.28: Northernhybridisierungen mit $\alpha 1(IV)$ - und $\alpha 2(IV)$ -spezifischen Proben .

Ca. 10 μ g Poly(A)+RNA verschiedener Herkunft wurden in einem Glyoxalgel aufgetrennt und durch Elektrotransfer auf Genescreen übertragen (Methoden, S. 144). Als Größenmarker wurde Hind III-verdaute λ -DNA eingesetzt.

Derselbe Filter wurde mit folgenden drei M13-Subklonen hybridisiert, die, wie im Methodenteil beschrieben (S.146), radioaktiv markiert worden waren:

- A. m15/1 ($\alpha 1$)
- B. mF4 ($\alpha 1$)
- C. ma2 ($\alpha 2$)

Die ermittelte Größe von 6.5 kb für die langste Variante der $\alpha 1$ -mRNA stimmt exakt überein mit dem Wert, der sich durch die Verfügbarkeit der mittlerweile fast vollständig bekannten cDNA-Sequenz von $\alpha 1$ (IV) ableiten läßt (Abb.29). Danach entfallen 5010 bp auf den codierenden Bereich, der somit 1670 Aminosäuren, einschließlich eines vermutlich 18 Aminosäuren langen N-terminalen Signalpeptids, bestimmt (Brazel, Dissertation 1987). Außerdem konnten 75 bp der nichttranslatierten Region am 5'-Ende aufgeklärt werden, die vermutlich ca. 150 bp lang ist (Brazel, Dissertation 1987). Nachdem sich die nichttranslatierte Region am 3'-Ende über 1278 bp erstreckt, ergibt die Summe dieser $\alpha 1$ -Sequenzen ca. 6440 bp. Da man annehmen kann, daß der Poly(A)-Schwanz nur 50 bis 60 bp lang ist (Mercer und Wake, 1985), erhält man für die längste Variante der $\alpha 1$ (IV)-mRNA einen Wert von 6.5 kb (Abb.29), in Übereinstimmung mit dem durch Northernhybridisierung ermittelten Wert.

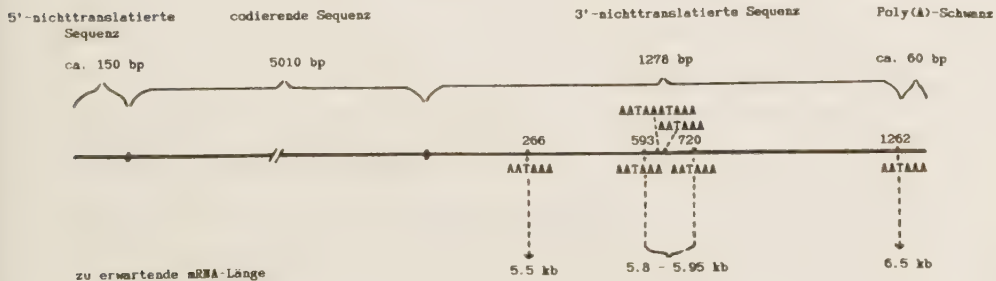


Abb.29: Positionen der AATAAA-Hexamere im 3'-nichttranslatierten Bereich von $\alpha 1$ (IV) von Maus
Die mRNA-Längen, die nach einem 3'-"processing" an den verschiedenen Stellen zu erwarten sind, wurden mitangegeben.

Für die 5.5 kb lange $\alpha 1(IV)$ -mRNA wurde angenommen, daß das 3'-"processing" dieser RNA im Bereich des ersten Polyadenylierungssignal (Pos.267 in Abb.25, S. 96) stattgefunden hatte. Folglich sollte die 5.5 kb-RNA bei Einsatz einer Hybridisierungsprobe, die 3' von der betreffenden AATAAA-Sequenz lokalisiert ist, nicht markiert werden. Um dies nachzuweisen, wurde eine zweite Northernhybridisierung mit demselben Filterstreifen und dem M13-Klon mF4 (Abb.22; S. 88) als Hybridisierungsprobe durchgeführt. Dieser Klon enthält den Abschnitt Pos.448 bis 689 in Abbildung 25, der 3' von der fraglichen Polyadenylierungsstelle liegt. Wie erwartet, war nach der Hybridisierung mit mF4 in der Spur mit der EHS-RNA kein positives Signal mehr im Bereich von 5.5 kb zu beobachten (Abb.28B). Dieses Resultat ist ein deutlicher Hinweis darauf, daß die 5.5 kb Bande eine RNA darstellt, die kurz nach dem ersten Poly(A)-Signal prozessiert wurde. Für die Benutzung dieses Signals sprechen zahlreiche weitere Befunde, die teilweise schon erläutert wurden. Die Charakterisierung des $\alpha 1$ -Klons pA14 (Teil III,2.1., S. 87) deutete auf den Signalcharakter des AATAAA-Hexamers in Pos.267 hin, da das 3'-Ende der cDNA dieses Klons in der betreffenden Region liegt. Zudem ist die Position dieses Hexamers und die flankierenden Sequenzen stark konserviert bei Maus und Mensch (Teil III,3., S. 99), einschließlich der Präsenz von potentiellen Signalsequenzen, die an der Auslösung des 3'-"processing" beteiligt sein sollen. Schließlich stimmt die beobachtete Länge von 5.5 kb für die fragliche mRNA sehr gut überein mit der von den Sequenzdaten abgeleiteten Länge (Abb.29).

Möglicherweise übt auch eines (oder mehrere) der AATAAA-Sequenzen zwischen Position 593 und 718 (Abb.29) eine Signalfunktion aus, wenn auch die Northernhybridisierungen von Abb.28 keinen Hinweis darauf geben. Bei humanen Fibroblasten wurde neben einer 6.5 kb- und einer 5.3 kb-Bande für $\alpha 1(IV)$ auch noch eine 5.9 kb-Bande beobachtet (Myers et al., 1986). Für die Entstehung dieser Variante sollte daher eine der beiden internen AATAAA-Hexamere (Pos. 603 und 713

in Abb.26) verantwortlich sein. Da die Positionen dieser beiden Hexamere bei Mensch und Maus konserviert sind, kann man annehmen, daß zumindest eines der beiden auch bei Maus eine Funktion ausübt.

Schließlich wurde mit dem Genescreenstreifen auch eine Hybridisierung mit einer $\alpha 2(IV)$ -spezifischen Probe durchgeführt, und zwar mit M13-Subklon ma2 (Abb.11, S. 42). Das Resultat zeigt Abbildung 28C. Die Position der markierten $\alpha 2$ -mRNA ist mit der Position der längsten $\alpha 1$ -mRNA (Abb.28A) deckungsgleich und hat somit auch eine Länge von 6.5 kb.

Wie bei $\alpha 1(IV)$ zeigt sich auch hier, daß die PYS-2-RNA offensichtlich den größten Anteil an Typ IV Kollagen-mRNA enthält. Nicht induzierte F9-Zellen exprimieren ebenfalls Typ IV Kollagen-mRNA, was auf spontane Differenzierung von einigen Zellen zurückzuführen ist (siehe Teil I,1., S. 17).

In der Spur mit HT-1080-RNA ist schwache Kreuzhybridisierung mit humaner $\alpha 2(IV)$ -RNA vorhanden, die sich, wie bei $\alpha 1(IV)$, in der Länge nicht nachweisbar von der Maus-RNA unterscheidet.

Überraschend war der Befund, daß die $\alpha 2(IV)$ -mRNA mit der längsten $\alpha 1(IV)$ -mRNA komigriert, also ebenfalls 6.5 kb lang ist. Für $\alpha 2(IV)$ ist zwar noch nicht die vollständige cDNA-Sequenz bekannt, man kann jedoch davon ausgehen, daß der codierende Bereich etwa die gleiche Länge hat wie bei $\alpha 1(IV)$. Da die nichttranslatierte Region um 330 bp kürzer ist als bei $\alpha 1(IV)$, war zu erwarten, daß die $\alpha 2$ -mRNA kürzer ist als die $\alpha 1$ -mRNA. Nachdem dies nicht der Fall ist, bleibt noch die Möglichkeit, daß eine im Vergleich zu $\alpha 1(IV)$ ausgedehntere nichttranslatierte Region am 5'-Ende die Länge der $\alpha 2$ -mRNA auf 6.5 kb anwachsen läßt. Demnach müßte für $\alpha 2(IV)$ eine 5'-nichttranslatierte Region von ca. 450 bp bis 500 bp gefordert werden. Dies wäre eine ungewöhnliche Länge für diesen Bereich - der Durchschnittswert liegt zwischen 20 bp und 80 bp (Kozak, 1984). Allerdings ist aus der Literatur bekannt, daß der nichtcodierende Abschnitt am 5'-Ende in Ausnahmefällen mehr als 500 bp erreichen kann (Kozak, 1984).

D. DISKUSSION

Typ IV Kollagen aus Basalmembranen war das erste Protein aus der Kollagenfamilie, für das auf molekularer und makromolekularer Ebene große strukturelle Unterschiede zu den interstitiellen Kollagenen Typ I bis III festgestellt wurden. Typ IV Kollagen ist nicht in Fibrillen organisiert, sondern bildet ein Netzwerk aus, das als Grundgerüst der Basalmembran fungiert. Die strukturellen Eigenschaften dieses Netzwerks müssen daher an die diversen biologischen Funktionen der Basalmembran angepaßt sein.

Voraussetzung für die Netzwerkbildung sind besondere Strukturelemente der Typ IV Kollagenmoleküle: die terminalen Domänen 7S (N-Terminus) und NC1 (C-Terminus) sind Quervernetzbereiche, deren spezifische Verknüpfung (vier Moleküle über die 7S-Domäne und zwei Moleküle über die NC1-Domäne) zur Ausbildung des Netzwerks führt. Der zwischen den Verknüpfungspunkten liegende tripelhelikale Bereich bestimmt die mechanische Stabilität und Flexibilität der entstandenen Gerüststruktur.

Im Rahmen der vorliegenden Arbeit wurde die Primärstruktur eines der beiden Quervernetzbereiche von Typ IV Kollagen, der NC1-Domäne am C-Terminus, für $\alpha 1(IV)$ und $\alpha 2(IV)$ vollständig aufgeklärt. Außerdem konnten die lückenhaften Peptidsequenzen des C-terminalen Bereichs der Tripelhelix ergänzt werden, insbesondere bei $\alpha 2(IV)$. Sowohl für die NC1-Domäne als auch für die Tripelhelix brachte die Analyse der Sequenzdaten und der Vergleich zwischen $\alpha 1(IV)$ und $\alpha 2(IV)$ wertvolle Informationen darüber, welche Strukturmerkmale für die besonderen Eigenschaften und Funktionen der NC1-Domäne einerseits und der Tripelhelix andererseits verantwortlich sind.

Die NC1-Domäne

Die Charakterisierung der NC1-Domäne ergab folgende Resultate, die hinsichtlich der strukturellen Bedeutung der NC1-Domäne als besonders wichtig erscheinen:

- a) Die Primärstruktur der globulären Domäne ist bei $\alpha 1(IV)$ und $\alpha 2(IV)$ stark konserviert (im Unterschied zum tripelhelikalen Abschnitt). Dies schließt die Positionen von jeweils zwölf Cysteinen mit ein.
- b) Der NC1-Bereich ist in beiden Ketten in zwei homologe Hälften unterteilt, wobei wiederum die Cysteine (je sechs pro Hälfte) homologe Positionen besetzen.
- c) Es bestehen Sequenzhomologien zu einem kurzen Abschnitt der Carboxypropeptide der interstitiellen Kollagene
- d) Der NC1-Bereich enthält keine Propeptidsequenzen, die vor der Aggregation abgespalten werden.
- e) Die DNA-Sequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$ sind kaum konserviert - der Selektionsdruck ist primär auf die Beibehaltung der Proteinstruktur gerichtet.

Die NC1-Domäne von Typ IV Kollagen und die Carboxypropeptide der interstitiellen Prokollagene sind in vieler Hinsicht vergleichbar. Beide sind als globuläre Domänen am C-Terminus der Kollagenmoleküle lokalisiert und annähernd gleich lang - ca. 250 Aminosäuren die Carboxypropeptide und ca. 230 Aminosäuren die NC1-Domäne. Auch hinsichtlich der Funktionen, die diesen C-terminalen Bereichen zugeordnet werden, gibt es wahrscheinlich zumindest teilweise Gemeinsamkeiten. Die Carboxypropeptide erfüllen eine wichtige Aufgabe bei der Zusammenlagerung von drei Peptidketten zu einem Kollagenmolekül (Fessler et al., 1981; Kühn, 1984;

Doege und Fessler, 1986). Es ist erwiesen, daß die monomeren Ketten der interstitiellen Kollagene zuerst im Bereich der Carboxypropeptide aggregieren, und von da ausgehend sich dann die Tripelhelix aufwindet. Zudem sind die Carboxypropeptide auch für die Kettenauswahl verantwortlich, aus der die spezifische Zusammensetzung der Moleküle resultiert (z.B. zwei $\alpha 1$ - und eine $\alpha 2$ -Kette bei Typ I Kollagen). Die gleichen Funktionen kann man auch für die NC1-Domäne annehmen. Ein entscheidender Unterschied zeigt sich jedoch bei der nachfolgenden Aggregation der Moleküle zu makromolekularen Strukturen. Die Carboxypropeptide werden vor der Fibrillenbildung abgespalten, was erst die Assoziation der Moleküle ermöglicht. Die NC1-Domäne dagegen verbleibt vollständig am Molekül und hat eine weitere wichtige Funktion, nämlich die Quervernetzung mit der NC1-Domäne eines zweiten Moleküls.

Trotz der zum Teil gemeinsamen Aufgaben, die den Carboxypropeptiden und der NC1-Domäne von Typ IV Kollagen zugeordnet werden können, und auch der vergleichbaren Strukturen, sind Sequenzhomologien zwischen den betreffenden Bereichen kaum vorhanden. Lediglich die Sequenz eines 33 Aminosäuren langen Abschnitts ist signifikant konserviert. Dieser Befund läßt vermuten, daß der konservierte Abschnitt an den gemeinsamen Funktionen, die den C-terminalen Domänen (Carboxypropeptide und NC1 Domäne) bei der Aggregation dreier α -Ketten zugeordnet werden, beteiligt ist. Dafür spricht auch, daß das betreffende Segment zu den am meisten konservierten Sequenzen der Carboxypropeptide gehört (Yamada et al., 1983). Bei den Carboxypropeptiden ist dieser 33 Aminosäuren lange Abschnitt jedoch nur einmal präsent, in der NC1-Domäne von Typ IV Kollagen dagegen zweimal. Dies ist die Folge der internen Homologie, die in den NC1-Sequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$ vorliegt.

Die Unterteilung der NC1-Region in zwei homologe Hälften war ein unerwarteter Befund. Vergleichbares tritt bei den interstitiellen Kollagenen nicht auf. Diese interne Homologie basiert möglicherweise auf der Duplikation eines Genabschnitts. Auf einen derartigen Vorgang könnten genomische Daten hinweisen, da sich in der Exon-Organisation eines Gens oft die Domänenstruktur des entsprechenden Proteins widerspiegelt (Gilbert, 1978). Die inzwischen teilweise bekannte Exon-Intron-Struktur des $\alpha 1(\text{IV})$ -Gens (Abb.30) läßt sich hiermit allerdings nicht korrelieren. Für $\alpha 1(\text{IV})$ von Mensch konnten zwei Exons des N-terminalen Bereichs der NC1-Sequenzen identifiziert und charakterisiert werden (Soininen et al., 1986). Das erste Exon, ein sogenanntes "junction"-Exon, enthält 71 bp der tripelhelikalen Region und 142 bp des NC1-Bereichs. Das zweite Exon, 178 bp lang, endet 13 bp vor dem Ende der ersten Subdomäne (333 bp lang). Die Trennung der NC1-Domäne in zwei homologen Hälften ist somit wahrscheinlich auf genomischer Ebene nicht verwirklicht. Offensichtlich haben die Prozesse, die zur Ausprägung der internen Homologie führten, vor oder unabhängig von der Exon-Intron-Organisation des Gens stattgefunden.

Hinsichtlich des Auftretens eines "junction"-Exons, das das Ende der Tripelhelix mit dem Beginn der C-terminalen Domäne vereint, gleicht die genomische Struktur von Typ IV Kollagen der der interstitiellen Kollagene (Abb.30). Allerdings ist das betreffende Exon bei $\alpha 1(\text{IV})$ kleiner als bei den interstitiellen Kollagenen. Das 3'-anschließende Intron dagegen ist mit 960 bp zumindest beim Menschen sehr viel länger als die entsprechenden Introns der interstitiellen Kollagene (Soininen et al., 1986).

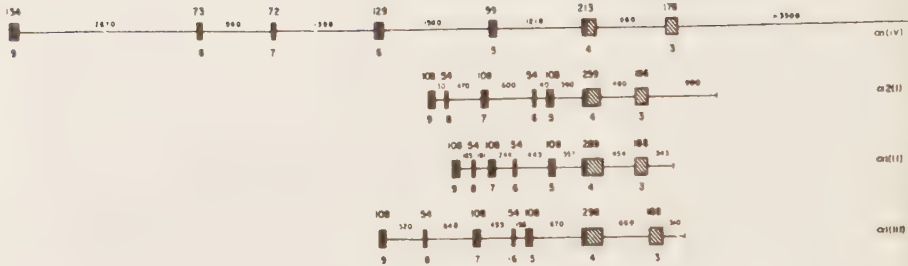


Abb.39: Exon-Intron-Struktur eines Genabschnitts des humanen $\alpha 1(IV)$ -Gens und Vergleich zu entsprechenden Bereichen der humanen $\alpha 2(I)$ -, $\alpha 1(II)$ - und $\alpha 1(IH)$ -Kollagen-Gene (aus: Soininen et al., 1986)
Bei den schwarzen Balken handelt es sich um Exons, die für tripelhelikale Abschnitte codieren. Die schraffierten Kästchen bezeichnen Exons, die NC1-Sequenzen von $\alpha 1(IV)$, bzw. Carboxypropeptid-Sequenzen der interstitiellen Kollagene enthalten. Die dazwischen liegenden Introns sind als Linie dargestellt. Die Zahlen oberhalb geben jeweils die Basenlänge der Exons und Introns an, die Zahlen unterhalb die Nummerierung der Exons. Die Nummerierung bei $\alpha 1(IV)$ wurde analog zu der der interstitiellen Kollagen-Gene durchgeführt.

Die interne Homologie des NC1-Bereichs schließt in beiden Ketten die Positionen der Cysteine (je sechs pro Halfte) mit ein. Die Konservierung der Cysteine und ihrer flankierenden Sequenzen sowohl zwischen $\alpha 1(IV)$ und $\alpha 2(IV)$, als auch in den homologen Halften einer einzelnen Kette, ist ein wichtiger Hinweis darauf, daß Disulfidbrücken eine entscheidende Rolle bei der Strukturbildung im NC1-Bereich spielen. Inzwischen liegen umfangreiche Informationen über die Lokalisation von Disulfidbrücken innerhalb der NC1-Domäne vor: bei der Analyse von isolierten NC1-Monomeren zeigte sich, daß Disulfidbrücken ausschließlich innerhalb der homologen Halften der einzelnen Ketten ausgebildet sind (Siebold, Dissertation 1986). Die beiden homologen Halften sind also nicht miteinander

quervernetzt, sondern stellen aufgrund interner Disulfidbrücken zwei klar abgegrenzte Subdomänen dar.

In den einzelnen Molekülen liegen auch keine Verknüpfungen zwischen Subdomänen verschiedener Ketten vor. Dies unterscheidet die NC1-Domäne von Typ IV Kollagen wiederum von den Carboxypropeptiden der interstitiellen Prokollagene, wo Disulfidbrücken zwischen den drei Ketten das Aggregat stabilisieren (Fessler et al., 1981).

Bei der Polymerisation von Typ IV Kollagen-Molekülen zur Netzwerkstruktur werden die internen Disulfidbrücken der Subdomänen zumindest teilweise gelöst und mit dem NC1-Bereich eines zweiten Moleküls durch Disulfidaustausch neu ausgebildet. Dafür sprechen die Analysen von isolierten NC1-Domänen aus quervernetztem Typ IV Kollagen. Nach Dissoziation dieser hexameren NC1-Domänen sind neben monomeren Formen zwei Arten von Dimeren nachweisbar (Weber et al., 1984). Diese Dimere bestehen jeweils nur aus $\alpha 1(IV)$ - oder nur aus $\alpha 2(IV)$ -Ketten (Siebold, Dissertation 1986). Die $\alpha 2(IV)$ -Dimere sind ausschließlich durch reduzierbare Bindungen verknüpft, Während man bei den $\alpha 1(IV)$ -Dimeren neben reduzierbaren Bindungen in geringem Umfang auch nichtreduzierbare Bindungen unbekannter Natur gefunden hat (Siebold, Dissertation 1986). Da nach Dissoziation der NC1-Hexamere außer Dimeren keine größeren kovalent verknüpften Aggregate beobachtet wurden, kann man annehmen, daß die räumliche Anordnung der einzelnen NC1-Domänen in einem hexameren Aggregat Quervernetzungen zwischen mehr als zwei Einheiten nicht zuläßt.

Die NC1-Sequenzen von $\alpha 1(IV)$ und $\alpha 2(IV)$ sind insgesamt hochgradig konserviert - über 60 % der Aminosäurepositionen sind identisch. Im Gegensatz dazu ist der anschließende tripelhelikale Bereich bei $\alpha 1(IV)$ und $\alpha 2(IV)$ kaum konserviert. Auf DNA-Ebene ist der hohe Verwandtschaftsgrad der NC1-Sequenzen allerdings weit-

aus weniger ausgeprägt, obwohl man davon ausgehen kann, daß beide Typ IV Gene einen gemeinsamen Ursprung haben. Eine Divergenz der beiden codierenden NC1-Sequenzen wurde offensichtlich insoweit toleriert, als die Konformation der Proteine nicht entscheidend verändert wurde. Dies deutet daraufhin, daß der Selektionsdruck der Evolution ausschließlich auf die Beibehaltung der Proteinstruktur dieses Bereichs gerichtet ist - ein Befund, der die funktionelle Bedeutung der NC1-Domäne für die Strukturbildung von Typ IV Kollagen unterstreicht.

Die Unterschiede in der Primärstruktur des NC1-Bereichs von $\alpha 1(IV)$ und $\alpha 2(IV)$ sind möglicherweise sogar notwendig, damit die spezifische Kettenauswahl bei der Assoziation von zwei $\alpha 1$ - und einer $\alpha 2$ -Kette zu einem Molekül stattfinden kann.

Die Tripelhelix

Durch die Analyse der $\alpha 2(IV)$ -spezifischen cDNA-Klone von Maus konnte die Proteinstruktur des über 500 Aminosäuren langen C-terminalen tripelhelikalen Abschnitts der $\alpha 2$ -Kette aufgeklärt werden. Auch für den korrespondierenden Bereich der $\alpha 1(IV)$ -Kette wurden teilweise die vorhandenen Lücken der Proteinstruktur geschlossen, teilweise wurden sie durch entsprechende Humansequenzen ergänzt. Es war somit für ein Drittel des tripelhelikalen Bereichs der Moleküle sowohl die Primärstruktur der $\alpha 1$ - als auch die der $\alpha 2$ -Kette bekannt, was zum ersten Mal eine eingehende Analyse der Typ IV Tripelhelix erlaubte.

Die Tripelhelix von Typ IV Kollagen weist gegenüber den Tripelhelices fibrillenbildender Kollagene strukturelle Besonderheiten auf. Auffällig ist vor allem, daß die tripelhelikale Sequenz $(Gly-X-Y)_n$ an zahlreichen Stellen durch nichthelikale Abschnitte (Abb.31) unterbrochen ist. Diese Unterbrechungen bewirken eine Auflockerung der tripelhelikalen Struktur. Dadurch sind bei Typ IV Kollagen die

fadenförmigen Moleküle nicht starr und geradlinig ausgerichtet wie bei den interstitiellen Kollagenen, sondern flexibel und anpassungsfähig. Für diese Flexibilität der Typ IV Tripelhelix sorgen vor allem die langen Unterbrechungen (z. B. I und VIII in Abb.31), die offensichtlich deutliche Konformationsänderungen verursachen. Sie sind im Gegensatz zu tripelhelikalen Abschnitten proteasesensitiv und außerdem in elektronenoptischen Studien als Knick im fadenförmigen Molekül sichtbar (Glanville et al., 1982; Hofmann et al., 1984). Möglicherweise fungieren diese langen Unterbrechungen wie Scharniere, die eine mehr oder weniger freie Drehbarkeit der Moleküle an der jeweiligen Stelle erlauben.

Die kurzen Unterbrechungen (z.B. II und VII in Abb.31) dagegen bewirken vermutlich keine Auflockerung der Struktur, sondern verursachen lediglich einen Knick des Moleküls. Dafür sprechen Untersuchungen des Serumfaktors Clq (Kilchherr et al., 1985). Dieses Protein enthält am N-Terminus einen tripelhelikalen Bereich, der von drei Peptidketten (A, B, und C) über eine Länge von 78 bis 81 Aminosäuren ausgebildet wird. Die ansonsten kontinuierliche Tripelhelix von Clq ist an einer Stelle unterbrochen, und zwar durch Insertion einer Aminosäure in Kette A (Pos.39) und durch Substitution eines Glycins durch ein Alanin in Kette C (Pos.36). Aus Energieminimierungsberechnungen und Modellbildung für dieses tripelhelikale Segment des Serumfaktors Clq ging hervor, daß die auftretende Unterbrechung nicht zu einer lokalen Destabilisierung der Tripelhelix führt, sondern lediglich zu einem Knick des Moleküls um ca. 60°. Eine vergleichbare Wirkung kann man daher für die kurzen tripelhelikalen Unterbrechungen bei Typ IV Kollagen annehmen, wenn auch hierfür noch kein experimenteller Nachweis vorliegt.

Die strukturelle Bedeutung der tripelhelikalen Unterbrechungen zeigt sich bei Homologievergleichen von $\alpha 1(IV)$ -Sequenzen verschiedener Spezies (Mensch, Maus, Rind). Die Unterbrechungen sind hier stark konserviert, und zwar hinsichtlich Anzahl, Position, Länge und Sequenz (Babel und Glanville, 1984). Offensichtlich

ist also ein Selektionsdruck auf die Beibehaltung dieser spezifischen Strukturelemente gerichtet. Möglicherweise wirken die tripelhelikalen Unterbrechungen nicht nur als destabilisierende Elemente, sondern haben noch weitere wichtige Funktionen, wie Interaktionen mit anderen Matrixkomponenten, z. B. Laminin, oder der Kontakt zu Zellen.

Beim Vergleich der tripelhelikalen Unterbrechungen von $\alpha 1(IV)$ und $\alpha 2(IV)$ sind dagegen kaum Übereinstimmungen festzustellen. Zumindest im C-terminalen Abschnitt der Tripelhelix ist eine Homologie nur insofern vorhanden, als beim Auftreten einer Unterbrechung in $\alpha 1(IV)$ stets auch eine Unterbrechung in entsprechender Position bei $\alpha 2(IV)$ vorliegt. Ansonsten ist weder die Länge noch die Sequenz dieser Abschnitte konserviert. $\alpha 2(IV)$ zeichnet sich dadurch aus, daß hier die Unterbrechungen fast durchwegs länger sind als bei $\alpha 1(IV)$, und daß außerdem zusätzliche Diskontinuitäten (II, V, VII, IX in Abb.31) auftreten. Aufgrund dieser Tatsache kann man die Auflockerung der tripelhelikalen Struktur von Typ IV Kollagen primär der $\alpha 2$ -Kette zuschreiben.

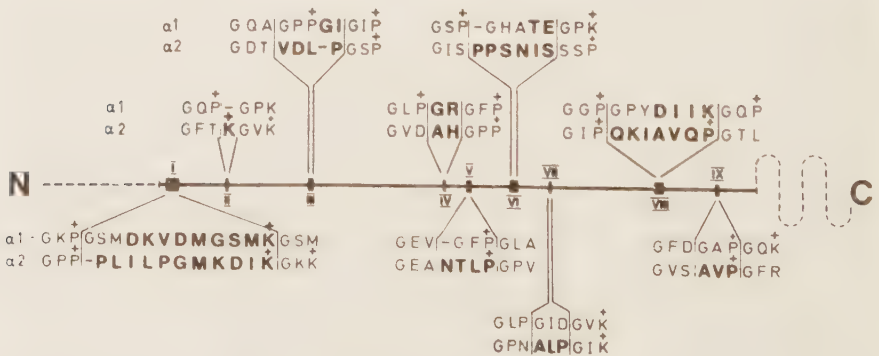


Abb.31: Unterbrechungen der tripelhelikalen Struktur im C-terminalen Abschnitt (510 Aminosäuren) der Tripelhelix von Typ IV Kollagen. An fünf Stellen (I, III, IV, VI und VIII) liegt in beiden α -Ketten eine Unterbrechung vor, an vier Stellen nur in der $\alpha 2$ -Kette (II, V, VII und IX). Die fett gedruckten Buchstaben kennzeichnen diejenigen Aminosäuren, die die kontinuierliche (Gly-X-Y)_n-Abfolge unterbrechen.

Trotz der Auflockerung der tripelhelikalen Struktur durch nichthelikale Abschnitte ist die thermische Stabilität der Typ IV Tripelhelix nicht reduziert - sie ist gegenüber den interstitiellen Kollagenen sogar leicht erhöht (Gelman et al., 1976). Die Tripelhelices der interstitiellen Kollagene werden allerdings durch die Fibrillenstruktur zusätzlich stabilisiert. Dieser Effekt entfällt für Typ IV Kollagen und wird möglicherweise durch die leicht erhöhte Stabilität der Moleküle kompensiert. Helixstabilisierende Einflüsse gehen vorwiegend von den Tripeptidkombinationen Gly-Pro-Hyp, Gly-Pro-Y und Gly-X-Hyp aus, wobei vor allem die Präsenz von Hydroxyprolin entscheidend ist (Sakibara et al., 1973; Berg und Prockop, 1973). Bei unterhydroxyliertem Kollagen liegt die Denaturierungstemperatur um mehr als 15 °C niedriger als bei normal hydroxyliertem Kollagen, so daß die Tripelhelix bei Körpertemperatur nicht mehr beständig ist. Die drei genannten Triplets sind bei den interstitiellen Kollagenen sehr häufig. Entsprechend hoch ist der Prolingehalt, wobei die Verteilung von Prolin auf X- und Y-Positionen ausgewogen ist: jeweils 30-35 % der X- und Y-Positionen sind durch Prolin besetzt. In den Y-Positionen liegt Prolin fast ausnahmslos in hydroxylierter Form vor (nur in der Y-Position wird Prolin modifiziert). Typ IV Kollagen hat zwar einen annähernd hohen Prolingehalt wie die interstitiellen Kollagene, jedoch ist die Verteilung dieser Aminosäure zwischen X- und Y-Positionen verschieden. Prolin ist bei Typ IV Kollagen in der X-Position deutlich unterrepräsentiert (16 % bei $\alpha 1(IV)$, 13 % bei $\alpha 2(IV)$), in der Y-Position jedoch sehr zahlreich vertreten (ca. 45 % bei $\alpha 1(IV)$, ca. 35 % bei $\alpha 2(IV)$). Da auch hier fast alle Proline in den Y-Positionen hydroxyliert sind, hat dies einen im Vergleich zu den interstitiellen Kollagenen signifikant höheren Hydroxyprolingehalt zur Folge.

Ein Großteil der Positionen dieser modifizierten Aminosäure ist außerdem bei $\alpha 1(IV)$ und $\alpha 2(IV)$ konserviert. Dies ist insofern bemerkenswert, weil ansonsten kaum Aminosäuren der X- und Y-Positionen identisch sind.

Aus diesen Daten geht somit hervor, daß die Stabilität der Typ IV Tripelhelix primär dem Hydroxyprolingehalt zuzuschreiben ist, der im Vergleich zu den interstitiellen Kollagenen überdurchschnittlich hoch ist. Dieser Stabilisierungseffekt geht in erster Linie von der $\alpha 1(IV)$ -Kette aus. $\alpha 2(IV)$ weist einen niedrigeren Hydroxyprolingehalt auf (vergleichbar mit dem der interstitiellen Kollagene) und ist zudem häufiger und durch längere nichthelikale Abschnitte unterbrochen. Man kann also feststellen, daß innerhalb eines Typ IV-Moleküls die beiden $\alpha 1$ -Ketten primär für die Stabilität der Tripelhelix verantwortlich sind, während die $\alpha 2(IV)$ -Komponente vor allem für die Auflockerung und Flexibilität sorgt.

Der hohe Grad posttranslationaler Modifikation der Aminosäure Lysin ist ebenfalls ungewöhnlich für kollagene Proteine. Lysin tritt in den beiden Ketten von Typ IV Kollagen mit wenigen Ausnahmen nur in der Y-Position der Triplets auf, wo es fast ausschließlich in hydroxylierter und glykosylierter Form (meistens Glukosyl-Galaktose) vorliegt. Ein beträchtlicher Anteil der Positionen dieser modifizierten Lysine ist wiederum bei $\alpha 1(IV)$ und $\alpha 2(IV)$ konserviert. Dies läßt darauf schließen, daß ihre Präsenz eine wichtige Rolle spielt. Die Funktion glykosylierter Hydroxylysine ist jedoch unklar, soweit sie nicht Quervernetzungen ausbilden wie z. B. im Bereich der 7S-Domäne von Typ IV Kollagen (Glanville et al., 1985). Bei den fibrillenbildenden Kollagenen wird eine Korrelation zwischen dem Gehalt an glykosylierten Hydroxylysinen und der Fibrillendicke beobachtet. Typ II Kollagen mit einem hohen Anteil glykosylierter Hydroxylysine bildet dünne Fibrillen aus, während das kaum glykosylierte Typ I Kollagen dicke Fibrillen ausbildet. Man vermutet, daß die Zuckerreste aus

sterischen Gründen die Zusammenlagerung dicker Fibrillen stören. Es wäre daher denkbar, daß die starke Glykosylierung bei Typ IV Kollagen die Funktion hat, die Aggregation der Tripelhelices zu verhindern.

Sequenzhomologien sind im tripelhelikalen Bereich zwischen $\alpha 1(IV)$ und $\alpha 2(IV)$ kaum vorhanden, im Gegensatz zur NC1-Domäne am C-Terminus. Ausgenommen sind die Positionen der modifizierten Aminosäuren Prolin und Lysin, sowie natürlich der Glycine. Diese beschränkte Homologie ist auch ein bemerkenswerter Unterschied zu den fibrillenbildenden Kollagenen. Hier sind die einzelnen α -Ketten stark konserviert, was im Zusammenhang mit dem hochgradig geordneten Prozess der Fibrillenbildung steht. Da Typ IV Kollagen eine andere makromolekulare Struktur ausbildet, wurde im Laufe der Evolution eine Divergenz der beiden α -Ketten toleriert, bzw. sogar gefördert. Dies hat dazu geführt, daß lediglich die Fähigkeit, eine stabile Tripelhelix auszubilden, konserviert wurde. Ausreichend hierfür ist offensichtlich, daß - mit wenigen Ausnahmen - ein Glycin in jeder dritten Aminosäureposition vorhanden ist, sowie ein hoher Gehalt an Hydroxyprolin.

Auch eine Unterteilung der Typ IV Tripelhelices in Subdomänen entsprechend den D-Domänen der interstitiellen Kollagene liegt nicht vor. Dies ist auch nicht erforderlich, da eine laterale Aggregation der Moleküle nicht stattfindet.

Bei den Kollagenen Typ I bis III verschiedener Spezies hatte die Aufklärung der genomischen Struktur gezeigt, daß die Exon-Intron-Organisation dieser Gene im Vergleich zu anderen Genen sehr ungewöhnlich ist. Vor allem aber führten die gewonnenen Informationen zu Spekulationen über den Verlauf der Evolution der Kollagen-Gene.

Die Exon-Intron-Struktur dieser Kollagen-Gene ist sehr ähnlich - Anzahl und Länge der Exons sind konserviert, lediglich die Längen der Introns variieren. Am

Beispiel des Gens von $\alpha 2(I)$ des Huhns (Boedtker et al., 1983) sollen im folgenden die charakteristischen Strukturmerkmale dieser Gene herausgestellt werden. Das $\alpha 2(I)$ -Gen umfaßt ca. 38 kb, wobei nur 10 % davon codierende Sequenzen sind. Dieser codierende Bereich ist in 51 Exons unterteilt, wovon 43 Exons auf die tripelhelikale Region entfallen, einschließlich zweier "junction"-Exons an C- und N-Terminus mit Anteilen nichthelikaler Sequenzen. Fast die Hälfte dieser tripelhelikalen Exons ist 54 bp lang und codiert für sechs vollständige Tripletts (Gly-X-Y). Darüberhinaus gibt es Exons mit Längen von 108 bp und 162 bp, die als Di- und Trimere der 54 bp-Einheit angesehen werden können. Außerdem treten Exons mit 45 bp und 99 bp auf, die jeweils durch Deletion von 9 bp in einem 54 bp und 108 bp Exon entstanden sein können. Dieses spezifische Muster der Exonlängen im tripelhelikalen Bereich hat zu der Auffassung geführt, daß ursprünglich ein singulärer Genabschnitt von 54 bp existiert hat, aus dem durch Amplifikation innerhalb von Intronsequenzen ein Urkollagen-Gen hervorgegangen ist. Dieses Ur-Gen sollte die gemeinsame Vorläuferform der heutigen Kollagen-Gene darstellen. Inzwischen muß man die aufgestellte Hypothese insoweit einschränken, als sie vermutlich nur für die fibrillenbildenden Kollagene zutrifft. Andere Kollagen-Gene wie das von *Drosophila melanogaster* (Monson et al., 1982), die von *Caenorhabditis elegans* (Kramer et al., 1982) und Typ IX Kollagen von Huhn (Lozano et al., 1985) weisen eine andere genomische Struktur auf. Bei Typ IX Kollagen hat keines der bisher analysierten Exons eine Länge von 54 bp oder 108 bp, wie es für die interstitiellen Kollagen-Gene charakteristisch ist. Die Untersuchungen von zwei Kollagen Genen von *Caenorhabditis elegans* ergaben, daß hier insgesamt nur ein bzw. zwei Introns enthalten sind, wobei keines innerhalb der tripelhelikalen Sequenzen lokalisiert ist. Bei *Drosophila melanogaster* wurde in einem für fast 500 Aminosäuren codierenden tripelhelikalen Abschnitt des Kollagen-Gens nur ein einziges Intron gefunden.

Mit großem Interesse wurden daher genomische Daten für die Tripelhelix von Typ IV Kollagen erwartet. Man mußte davon ausgehen, daß zumindest für die Abschnitte, die tripelhelikale Unterbrechungen enthalten, andere Exonlängen auftreten als bei den interstitiellen Kollagenen. Mittlerweile liegen genomische Daten für Teile der $\alpha 1(IV)$ -Tripelhelix von Mensch und Maus (Sakurai et al., 1985; Soininen et al., 1986) und der $\alpha 2(IV)$ -Tripelhelix von Maus vor (Kurkinen et al., 1985). Bei $\alpha 1(IV)$ wurden Exonlängen von 72, 73, 81, 99, 129, 134 und 178 bp gefunden (siehe auch Abb.30), für $\alpha 2(IV)$ 64, 123 und 182 bp. Diese Exongrößen lassen sich nicht von einer 54 bp-Einheit ableiten - nicht einmal diejenigen Exons, die für kontinuierliche tripelhelikale Abschnitte ohne Unterbrechungen codieren. Ungewöhnlich ist außerdem, daß fast alle bisher bekannten Exons von $\alpha 1(IV)$ und $\alpha 2(IV)$ erst mit der zweiten Base eines Glycin-Codons beginnen. Bei den Genen der fibrillenbildenden Kollagene steht am 5'-Ende eines Exons stets ein vollständiges Glycin-Codon. Diese Befunde deuten daraufhin, daß die Evolution von Typ IV Kollagen unabhängig von der der interstitiellen Kollagen stattgefunden hat.

Ein Homologievergleich von $\alpha 1(IV)$ und $\alpha 2(IV)$ hinsichtlich der Genstruktur ist zum gegenwärtigen Zeitpunkt noch nicht möglich, da die analysierten Genabschnitte in verschiedenen Bereichen liegen.

Die Introngrößen von Typ IV Kollagen (Abb.30) übersteigen, soweit bisher ersichtlich, deutlich die der fibrillenbildenden Kollagen-Gene (Soininen et al., 1986). Somit kann man davon ausgehen, daß die Typ IV Gene insgesamt extrem lang sind (60 bis 70 kb, Yamada, pers. Mitteilung).

Die nichttranslatierten Regionen am 3'-Ende

Die Aufklärung der Primärstruktur der nichttranslatierten Region von $\alpha 1(IV)$ und $\alpha 2(IV)$ von Maus, sowie die Charakterisierung der mRNA ergaben folgende Resultate:

- a) Die 3'-nichttranslatierten Regionen sind ungewöhnlich lang, 1278 bp bei $\alpha 1(IV)$ und 969 bp bei $\alpha 2(IV)$.
- b) Der 3'-Bereich von $\alpha 1(IV)$ enthält sieben kanonische Polyadenylierungssignale (AATAAA), der von $\alpha 2(IV)$ nur ein einziges.
- c) Für $\alpha 1(IV)$ treten verschieden lange mRNAs auf, was sich mit der Lokalisation von AATAAA-Signalsequenzen im 3'-Bereich korrelieren läßt.
- d) Das Muster der Längenheterogenität variiert abhängig vom Zelltyp und möglicherweise von der Wachstumsphase der Zellen.
- e) Bei $\alpha 1(IV)$ ist eine starke Interspezieshomologie (Mensch/Maus) ausgeprägt, die die Positionen von vier der sieben AATAAA-Sequenzen der Maus - wahrscheinlich derjenigen mit Signalwirkung - miteinschließt.
- f) Der Sequenzvergleich von $\alpha 1(IV)$ und $\alpha 2(IV)$ zeigt so gut wie keine Homologien.

Aus den aufgeführten Fakten geht hervor, daß die 3'-nichtcodierende Region - zumindest bei $\alpha 1(IV)$ - einen wichtigen Genabschnitt darstellt. Dafür sprechen vor allem die Konservierung der Struktur dieses Bereichs bei verschiedenen Spezies, sowie die Längenheterogenität der mRNA, die abhängig vom Zelltyp variieren kann.

Inzwischen werden zunehmend Gene bekannt, wo die Längenverteilung der mRNA-Population je nach Entwicklungsphase und/oder Gewebeart unterschiedlich ausgeprägt ist, basierend auf der alternativen Benutzung multipler Polyadenylierungssignale. Als Beispiele seien das Alkoholdehydrogenase-Gen von Mensch

(Bilanchone et al., 1986), das Tropomyosin-Gen von Drosophila (Boardman et al., 1985) und das Vimentin-Gen von Huhn (Capetanaki et al., 1983) genannt. Auch bei den Kollagenen wurde beobachtet, daß die Längenheterogenität der mRNA abhängig vom Zelltyp variieren kann. Bei $\alpha 1(I)$ von Mensch z. B. zeigt sich in Fibroblasten ein anderes Muster als in der Plazenta (Barsh et al., 1984).

Es stellt sich somit die Frage, welchen Einfluß ein alternatives 3'-"processing" auf Struktur und Funktion der mRNA hat.

Im diesem Zusammenhang sind experimentelle Daten aus der Arbeit von Pfarr et al. (1986) sehr aufschlußreich. Hier wurden die 3'-Regionen verschiedener Gene (Typ I Kollagen, frühes SV40 Gen und Rinder-Wachstumsfaktor) jeweils hinter ein und dasselbe Indikatorgen (Galaktokinase mit SV40-Promotor) geschaltet. Diese Konstrukte wurden dann in eukaryontischen Zellen zur Expression gebracht. Es zeigte sich, daß die Effizienz der Expression stark differierte: die nicht-translatierte Region von Typ I Kollagen (Mensch) bewirkte die gleiche Expressionsrate wie die des frühen SV40-Gens; als dreimal effektiver erwies sich dagegen der 3'-Bereich des Rinder-Wachstumsfaktors. Die unterschiedliche Ausbeute an Genprodukten beruhte darauf, daß das Niveau der jeweiligen mRNA-Konzentration variierte. Als Ursache hierfür wird angesehen, daß, abhängig von der Art der nichtcodierenden Region, die Bildung von prozessierter mRNA mehr oder weniger schnell erfolgt oder/und die Stabilität der mRNA verändert wird. Diese Versuche zeigen, daß die 3'-nichtcodierende Region der mRNA für den "steady-state level" der mRNA-Konzentration verantwortlich sein kann, und damit einen wichtigen Faktor bei der Regulation der Genexpression darstellt.

Bei Typ I Kollagen gibt es Hinweise darauf, daß das Hormon Dexamethason in Fibroblasten speziell die Stabilität der Typ I-mRNA vermindert (Raghow et al., 1986). Interferon hemmt in rheumatoiden Synovialzellen die Typ I und Typ III Kollagen Synthese, möglicherweise ebenfalls durch Reduktion der RNA-Stabilität

(Stephenson et al., 1985). Ob diese Einflüsse durch Veränderungen der nicht-translatierten Regionen der mRNA vermittelt werden, ist nicht geklärt.

Es gibt auch Indizien dafür, daß der 3'-nichtcodierende Bereich der mRNA auf eine andere Weise in die Regulation der Genexpression involviert sein kann, und zwar durch den Einfluß auf die Translatierbarkeit der RNA. Darauf deuten experimentelle Daten aus der Analyse von Typ I Kollagen (Focht und Adams, 1984) und auch Typ IV Kollagen (Laurent et al., eingereicht) hin. Typ I Kollagen wird von differenzierten Chondrozyten nicht synthetisiert. Überraschenderweise hat man jedoch Typ I-spezifische mRNA in diesen Zellen gefunden. Die betreffende mRNA-Population weist andere Längen auf als die aus Typ I Kollagen-synthetisierenden Zellen wie z. B. Haut- und Sehnenfibroblasten. Ob das veränderte Längenmuster ausschließlich auf der Verwendung verschiedener Polyadenylierungssignale beruht, ist noch nicht geklärt. Jedenfalls wird Typ I-mRNA aus Chondrozyten weder in vivo, noch in vitro nachweisbar translatiert. Dieser Mangel an Translatierbarkeit in Chondrozyten konnte demnach durch den gegenüber anderen Zelltypen vermutlich veränderten 3'-nichtcodierenden Bereich der mRNA verursacht werden.

Auch bei Typ IV Kollagen wurde ein vergleichbarer Effekt beobachtet. $\alpha 1(IV)$ -mRNA aus dem EHS-Maustumor wurde durch einen Sucrose-Gradienten nach Größe fraktioniert. Zellfreie Translationen der einzelnen Fraktionen zeigten, daß die kürzeste RNA-Spezies von 5.5 kb effizienter translatiert wird als die längste von 6.5 kb (Laurent et al., eingereicht). Die kurze Variante der $\alpha 1$ -mRNA ist im EHS-Maustumor im Vergleich zu anderen Zelltypen relativ häufig (siehe Abb.28, S. 106). Dies würde bedeuten, daß die $\alpha 1(IV)$ -mRNA-Population aus EHS-Tumorzellen insgesamt effizienter translatiert wurde als z. B. die aus PYS-2 oder F9 Zellen. Inwieweit sich eine unterschiedliche Translatierbarkeit der $\alpha 1(IV)$ -mRNA auf die Biosyntheserate von Typ IV Kollagen auswirkt, ist unklar, vor allem auch des-

halb, weil ein derartiger Effekt auf $\alpha 1(IV)$ beschränkt sein dürfte. Für $\alpha 2(IV)$ entfällt nämlich der mögliche Einfluß von Regulationsvorgängen, der mit dem Auftreten von multiplen Signalelementen gegeben ist. Wenn $\alpha 1(IV)$ und $\alpha 2(IV)$ post-transkriptional einer unterschiedlichen Regulation unterworfen sind, dann stellt sich die Frage, wie sich dies auf die Syntheserate von Typ IV Kollagen-Molekülen auswirkt. In diesem Zusammenhang wäre es von Interesse, in verschiedenen Zelltypen das stöchiometrische Verhältnis der einzelnen mRNA-Populationen zueinander und die jeweilige Translatierbarkeit zu bestimmen. Wegen der spezifischen Zusammensetzung der Typ IV Kollagen Moleküle (zwei $\alpha 1$ - und eine $\alpha 2$ -Kette) ist ein Überschuß von $\alpha 1$ -Ketten erforderlich, was auch bei Analysen von Zellextrakten und Zellkulturmedien beobachtet wird (Fessler und Fessler, 1982). Bisher ist noch nicht bekannt, ob und wie die Koordination der Expression von $\alpha 1(IV)$ und $\alpha 2(IV)$ stattfindet.

Alternatives 3'-"processing" kann auch auf andere Weise die Genexpression regulieren. Bisher sind drei Fälle bekannt, wo von der Wahl definierter Polyadenylierungssignale das Muster von differentiellem Exon-Intron-"splicing" abhängt. Es handelt sich hierbei um die späte Transkriptionseinheit des Adenovirus (Nevins, 1983), das Calcitonin-Gen (Amara et al., 1984) und das Gen der schweren Kette der IgM Immunglobuline (Mather et al., 1984). Beim Calcitonin/CGRP-Gen ("Calcitonin gene-related peptide") z. B. findet geweberspezifisch eine unterschiedliche Selektion zwischen zwei Polyadenylierungssignalen statt. Von der Auswahl zwischen den beiden Signalen hängt ab, welche Variante des anschließenden differentiellen "splicing" ausgelöst wird, was in einem Fall zum Calcitonin als Genprodukt, im anderen zum CGRP führt. Bei diesen Untersuchungen zeigte sich, daß die Prozessierung des 3'-Endes der mRNA mit hoher Wahrscheinlichkeit noch vor dem Exon-Intron-"splicing" erfolgt. Außerdem

deuten die Analysen daraufhin, daß stets ein vollständiges Primärtranskript gebildet wird, dessen alternatives 3'-"processing" dann post-transkriptional gesteuert wird. Im Unterschied dazu entscheidet beim Adenovirus vermutlich die Termination der Transkription darüber, welche der Polyadenylierungsstellen benutzt wird - die Auswahl steht hier also unter Transkriptionskontrolle. Bei Typ IV Kollagen gibt es bisher keinen Hinweis auf differentielles "splicing". Die Präsenz multipler Polyadenylierungssignale dürfte also in diesem Zusammenhang keine Rolle spielen.

Mit Ausnahme des EHS-Tumors findet in allen bisher untersuchten Zelltypen das 3'-"processing" der $\alpha 1(IV)$ -mRNA bevorzugt nach dem 3'-gelegenen Polyadenylierungssignal statt - die längste mRNA-Spezies (6.5 kb) von $\alpha 1(IV)$ ist weitaus am häufigsten, die kleinste (5.5 kb) oft am zweithäufigsten. Dieses Muster unterscheidet sich von dem der interstitiellen Kollagene, wo in keinem Fall die längste Variante überwiegt.

Die Dominanz der 6.5 kb-RNA bei $\alpha 1(IV)$ kann verschiedene Ursachen haben. Sie könnte allein darauf beruhen, daß diese RNA-Spezies durch die Struktur ihrer nichttranslatierten Region in besonderem Maße gegen Degradation geschützt ist. Dagegen spricht jedoch, daß die gleiche RNA in EHS-Tumorzellen weniger prominent ist (siehe Abb.28, S. 106). Demzufolge kann man annehmen, daß ein Selektionsvorgang involviert ist, der in den einzelnen Zelltypen mehr oder weniger wirksam ist. Die Selektion der 3'-terminalen Polyadenylierungsstelle konnte entweder auf Transkriptionsebene ablaufen wie beim Adenovirus (Nevins, 1983), oder aber post-transkriptional wie beim Calcitonin-Gen (Amara et al., 1984). Im zweiten Fall kann man davon ausgehen, daß ein besonders starker Selektionsmechanismus involviert sein muß, bzw. daß die Signalwirkung der anderen potentiellen Polyadenylierungsstellen weitgehend ausgeschaltet sein muß.

Wenn nämlich die Möglichkeit gegeben wäre, daß die Spaltung nach den Polyadenylierungssignalen statistisch verteilt erfolgte, und ein und dasselbe Primärtranskript gleichzeitig an verschiedenen Stellen geschnitten und polyadenyliert würde, dann würde dies die Bildung der kürzeren RNA-Spezies begünstigen. Aufgrunddessen kann man annehmen, daß die 3'-terminale Polyadenylierungsstelle bei $\alpha 1(IV)$ durch cis-aktive Elemente einen außergewöhnlich starken Signalcharakter hat, oder/und die Einwirkung trans-aktiver Faktoren zu einer bevorzugten Benutzung führt. Für die Beteiligung trans-aktiver Faktoren spricht, daß die Selektion der 3'-terminalen Polyadenylierungsstelle deutlich reduziert sein kann wie im EHS-Tumor, wo die 6.5 kb-RNA und die 5.5 kb-RNA in ungefähr äquimolaren Mengen gebildet werden.

Beim Vergleich der $\alpha 1(IV)$ -mRNA-Populationen von Maus und Mensch zeigte sich, daß auch in den bisher analysierten humanen Zelltypen die längste RNA (6.5 kb) bei weitem überwiegt. In humanen Fibroblasten sind zusätzlich eine 5.9 kb- und eine 5.3 kb-RNA nachweisbar (Brinker et al., 1985), in HT-1080 Zellen fehlt vermutlich die mittlere 5.9 kb-RNA (Pihlajaniemi et al., 1985). Unterschiedliche Verteilungen der verschieden langen $\alpha 1(IV)$ -mRNAs treten also auch beim Menschen auf. Eine insgesamt ähnliche Längenverteilung der $\alpha 1(IV)$ -mRNA bei Mensch und Maus war zu erwarten, nachdem Sequenzhomologien der 3'-nichttranslatierten Regionen in hohem Ausmaß bestehen, einschließlich der Positionen von vier AATAAA-Hexameren. Aufgrund der Konservierung dieser vier Hexamere kann man davon ausgehen, daß alle eine Signalfunktion haben; für das 5'- und das 3'-terminale AATAAA ist dies nachgewiesen (Bildung der 6.5 kb- und der 5.5 kb-RNA), für die beiden mittleren ist nicht klar, ob nur eines oder beide ein 3'-"processing" auslösen können.

Die hochgradige Homologie der 3'-Sequenzen von $\alpha 1(IV)$ bei Mensch und Maus ist jedenfalls ein weiteres Indiz dafür, daß diese Struktur eine wichtige Aufgabe erfüllt. Interspezieshomologie von 3'-nichtcodierenden Regionen von Vertretern einer Genfamilie ist schon in mehreren Fällen nachgewiesen, z. B. bei den Aktin-Genen (Yaffe et al., 1985). Die 3'-Regionen des Aktins aus Skelettmuskel z. B. sind stark konserviert bei verschiedenen Organismen. Andererseits zeigte sich, daß innerhalb einer Genfamilie die entsprechenden Bereiche nicht ähnlich sind: bei den Aktinen aus Skelett- und aus Herzmuskelzellen besteht keine Ähnlichkeit der 3'-Sequenzen. Nach dem gleichen Schema verhalten sich auch die 3'-nicht-translatierten Bereiche von $\alpha 1(IV)$ und $\alpha 2(IV)$. $\alpha 2(IV)$ weist hier eine völlig andere Struktur auf als $\alpha 1(IV)$. Die bei $\alpha 1(IV)$ konservierten Strukturelemente spielen offensichtlich für die Expression von $\alpha 2(IV)$ keine Rolle. Dies wirft, wie bereits erwähnt, die Frage auf, ob und inwieweit hier eine koordinierte Regulation der Genexpression von $\alpha 1(IV)$ und $\alpha 2(IV)$ stattfindet.

Kurzer Ausblick

Mittlerweile sind auch für den N-terminalen Bereich der beiden α -Ketten von Typ IV Kollagen die Peptidsequenzen fast vollständig bekannt (Siebold, Dissertation 1986; Brazel, Dissertation 1987). Somit kann die Charakterisierung von Typ IV Kollagen auf Proteinebene demnächst abgeschlossen werden.

Der Schwerpunkt der Arbeiten wird sich daher auf die Aufklärung der genomischen Struktur verlagern. Dies stellt ein aufwendiges Projekt dar in Anbetracht der zu erwartenden Größe der Gene und der Vielzahl von Exons bzw. Introns. Durch die Verfügbarkeit der fast vollständigen cDNA-Sequenzen sind hierfür jedoch die besten Voraussetzungen gegeben. Von großem Interesse ist der Vergleich der genomischen Strukturen von $\alpha 1(IV)$ und $\alpha 2(IV)$ im Hinblick darauf, ob die Exon-Intron-Verteilung homolog ist und damit eine enge Verwandtschaft der Typ IV-Gene besteht.

Außerdem sollen natürlich die Promotorbereiche der Typ IV-Gene analysiert werden, um Informationen über die Transkriptionskontrolle zu erhalten.

Schließlich sind, wie aus der Diskussion hervorgeht, noch viele Fragen hinsichtlich der Funktionen der 3'-nichttranslatierten Regionen offen, deren Klärung im Zusammenhang mit der Regulation der Genexpression bei Typ IV Kollagen von Interesse ist.

E. ZUSAMMENFASSUNG

Durch die Isolierung und Analyse von cDNA-Klonen konnte die Primärstruktur der NC1-Domäne am C-Terminus von Typ IV Kollagen, sowie des C-terminalen Bereichs der Tripelhelix aufgeklärt werden. Ferner wurden die 3'-nichtcodierenden Regionen von $\alpha 1(IV)$ und $\alpha 2(IV)$ und die jeweiligen mRNA-Populationen charakterisiert.

Die NC1-Domäne

Die NC1-Region umfaßt bei $\alpha 1(IV)$ 229 Aminosäuren, bei $\alpha 2(IV)$ 227 Aminosäuren. Beide Ketten sind in diesem Bereich sehr ähnlich - über 60 % der Aminosäuren in korrespondierenden Positionen sind identisch. Dies schließt die Positionen von zwölf Cysteinen mit ein, die hinsichtlich der Stabilisierung von Sekundär- und Tertiärstrukturen sehr wichtig sind.

Beide Peptidketten sind im Bereich der NC1-Domäne in zwei homologe Hälften unterteilt, wo neben zahlreichen identischen Aminosäuren auch die Cysteine (je sechs pro Hälfte) korrespondierende Positionen besetzen. Jede dieser Subdomänen bei $\alpha 1(IV)$ und $\alpha 2(IV)$ zeigt Homologien zu einem kurzen Abschnitt der Carboxypropeptide der interstitiellen Kollagene. Diese konservierte Sequenz erfüllt möglicherweise eine wichtige Funktion bei der Assoziation von drei α -Peptidketten und der Kettenauswahl.

Bei der Aggregation von Typ IV-Molekülen zur Netzwerkstruktur werden keine C-terminalen Propeptide abgespalten wie bei den interstitiellen Kollagenen.

Die Tripelhelix

Für $\alpha 1(IV)$ wurden durch die Sequenzierung eines 495 bp langen cDNA-Abschnitts die lückenhaften Peptidsequenzen des C-terminalen Bereichs der Tripelhelix ergänzt. Für $\alpha 2(IV)$ wurde die Primärstruktur der 511 Aminosäuren langen C-terminalen Region der Tripelhelix aufgeklärt. Dadurch war es erstmals möglich,

korrespondierende Abschnitte (511 Aminosäuren) von $\alpha 1(IV)$ und $\alpha 2(IV)$ zuzuordnen und zu vergleichen. Die tripelhelikale Struktur ist in dem betreffenden Bereich neunmal unterbrochen: an fünf Positionen in beiden Ketten, wobei die Unterbrechungen in $\alpha 2(IV)$ meistens länger sind. Sequenzhomologien zwischen $\alpha 1(IV)$ und $\alpha 2(IV)$ innerhalb der Unterbrechungen bestehen nicht. In der $\alpha 2$ -Kette sind außerdem zusätzlich vier weitere Unterbrechungen vorhanden.

Neben dem periodischen Auftreten von Glycin gibt es kaum Sequenzhomologien zwischen $\alpha 1(IV)$ und $\alpha 2(IV)$ - lediglich die Positionen eines Großteils der modifizierten Aminosäuren Hydroxyprolin und Glykosyl-Hydroxylysin sind konserviert.

Die 3'-nichtcodierenden Regionen

Dieser Bereich umfaßt bei $\alpha 1(IV)$ 1278 bp, bei $\alpha 2(IV)$ 969 bp. Bei $\alpha 1(IV)$ treten in der 3'-nichttranslatierten Region sieben Polyadenylierungssignale (AATAAA) auf, bei $\alpha 2(IV)$ nur ein einziges. Auch hinsichtlich der Sequenzen sind die beiden 3'-Bereiche nicht ähnlich, dagegen besteht bei $\alpha 1(IV)$ große Interspezieshomologie zwischen Mensch und Maus.

Bei $\alpha 1(IV)$ ist eine Längenheterogenität der mRNA (6.5 kb und 5.5 kb) ausgeprägt, die auf der Benutzung verschiedener Polyadenylierungssignale beruht und abhängig vom Zelltyp variiert. Die $\alpha 2$ -spezifische mRNA ist ebenfalls 6.5 kb lang.

Die Analyse der cDNA dreier Genabschnitte - des NC1-Bereichs, des C-terminalen tripelhelikalen Bereichs und der 3'-nichtcodierenden Region - zeigte, daß $\alpha 1(IV)$ und $\alpha 2(IV)$ auf DNA-Ebene kaum ähnlich sind. Die codierenden Abschnitte sind nur insoweit konserviert, als die Beibehaltung der Peptidstrukturen dies erfordert. Im nichttranslatierten Bereich ist keinerlei Ähnlichkeit mehr zwischen $\alpha 1(IV)$ und $\alpha 2(IV)$ vorhanden, was auf differierende Funktionen dieser Regionen hindeutet.

F. MATERIAL UND METHODEN

I. MATERIAL

1. Verwendete Substanzen und Bezugsquellen

Acrylamid	Roth (Karlsruhe), Biomol (Ilvesheim)
Agarose	Seakem (Rockland, USA)
Amberlite	Serva (Heidelberg)
Ammoniumpersulfat	BDH (Poole, UK)
Ampicillin	Sigma (München)
Bacto-Agar	Difco (Detroit, USA)
"Bind"-Silan	LKB (München)
Chloramphenicol	Sigma (München)
Dimethylformamid	Fluka AG (Neu-Ulm)
DMS	EGA-Chemie (Steinheim)
Ficoll	Serva (Heidelberg)
Genescreen	NEN (Dreieich)
Glyoxal	Sigma (München)
Guanidiniumhydrochlorid	BRL (Bethesda, USA)
Hefeextrakt	Difco (Detroit, USA)
Hydrazin	EGA-Chemie (Steinheim)
IPTG	Biomol (Ilvesheim)
Mercaptoäthanol	Roth (Karlsruhe)
Methylquecksilberhydroxid	Ventron (Karlsruhe)
Nitrozellulosefilter	Schleicher & Schuell (Dassel)
N,N'-Methylenbisacrylamid	Roth (Karlsruhe)
Oligo(dT)-Zellulose	P-L Pharmacia (Freiburg)
PEG 6000	Serva (Heidelberg)
Phenol	BRL (Bethesda, USA)
Piperidin	EGA-Chemie (Steinheim)
Rinderserumalbumin	Sigma (München), Boehringer (Mannheim)
Sarcosyl	Ciba Geigy (Basel, Schweiz), Serva (Heidelberg)
SDS	Serva (Heidelberg)
Sephacryl S-200, S-300, S-400,	
Sephadex G-50	Pharmacia (Freiburg)
TeMED	Serva (Heidelberg)

Tetracyclin	Sigma (München)
Trypton	Difco (Detroit, USA)
X-Gal	Biomol (Ilvesheim), Bachem (Schweiz)

Alle anderen Chemikalien wurden von der Firma Merck (Darmstadt) bezogen. Für Autoradiogramme wurden Filme (XAR-5, Ortho-G und XS-20) der Firma Kodak verwendet.

Radiochemikalien

^{32}P - γ -ATP (3000 Ci/mmol) wurde von Amersham (Braunschweig) bezogen, ^{32}P - α -dNTPs (3000 Ci/mmol) von Amersham und NEB (Dreieich), ^{35}S - γ -ATP (500 Ci/mmol) von NEB.

Enzyme

Die verwendeten Restriktionsenzyme stammten fast ausschließlich von der Firma Boehringer (Mannheim), ansonsten von Biolabs (Schwalbach). Ligase, alkalische Phosphatase, DNA-Polymerase (Kornberg und Klenow), DNase und RNase wurden ebenfalls von Boehringer bezogen, Lysozym von Sigma, Polynukleotidkinase von P-L Pharmacia und Proteinase K von Merck.

Nukleotide und Nukleinsäuren

Nukleotide (dNTPs, ddNTPs, ATP) wurden von der Firma P-L Pharmacia bezogen. Die eingesetzten Sequenzierungs- und Hybridisierungsprimer, sowie Heringssperma-DNA stammten von Boehringer, Hind III-geschnittene λ -DNA von BRL, Hae II-geschnittene ϕ x-DNA von Biolabs, die Vektoren pUC8 und pUC9 von P-L Pharmacia und die Vektoren M13tg130 und tg131 von Amersham.

2. E. coli-Stämme

C600.4	F ⁻ , thi1, thr1, leuB6, lacY1, supE44, tonA21, (Maniatis et al, 1982)
DH1	F ⁻ , recA1, endA1, gyrA96, thi, hsdR17, supE44, relA1, (Hanahan, 1983)
DH5	Abkömmling des DH1 (Hanahan, 1985)
JM83	ara, Δ (lac-proAB), rspL(=stra), ϕ 80, lacZ Δ M15 (Vieria und Messing, 1982)
JM101	supE, thi, Δ (lac-proAB), (F', traD36, proAB, lacI Δ Z Δ M15) (Yanisch-Perron et al., 1985)
JM109	recA1, endA1, gyrA96, thi, hsdR17, supE44, relA1, Δ (lac-proAB), (F', traD36, proAB, lacI Δ Z Δ M15) (Yanisch-Perron et al., 1985)

3. Nährmedien

Vollmedien (Angaben pro Liter):

LB	Trypton	10 g
	Hefeextrakt	5 g
	NaCl	10 g
YT	Trypton	8 g
	Hefeextrakt	5 g
	NaCl	5 g
TYM	Trypton	2 g
	Hefeextrakt	0.5 g
	NaCl	6 g
	MgCl ₂	1 g

Zur Herstellung von Platten wurde den Medien jeweils 15g/l Bacto-Agar zugesetzt. Der bei der Transformation mit M13 verwendete Weichagar enthielt 0.6 % Agarose in YT-Medium.

Minimalmedium (Angaben pro Liter):

Bacto-Agar 15 g

M9 Salze:

Na₂HPO₄ 6 g

KH₂PO₄ 3 g

NaCl 0.5 g

NH₄Cl 1 g

Der Agar und die M9-Salze wurden getrennt autoklaviert. Nach dem Autoklavieren werden folgende, steril filtrierte Lösungen zugegeben:

1 M MgSO₄ 1 ml

0.1 M CaCl₂ 1 ml

1 % Thiamin-HCl 1 ml

40 % Glukose 5 ml

4. Häufig verwendete Puffer und Lösungen

TBE: 90 mM Tris
 90 mM Borsäure
 2.5 mM EDTA

SSC: 0.15 M NaCl
 0.015 M Na-Citrat
 pH 7

TE: 10 mM Tris-HCl/pH 7.6
 1 mM EDTA

Auftragspuffer

für nicht-denaturierende Gele (5xLösung):

25 % Glycerin
2.5 % SDS
0.25 % Farbstoff (Bromphenolblau und Xylencyanol)

für Sequenzgele:

10 mM NaOH
1 mM EDTA
0.05 % Farbstoff (Bromphenolblau und Xylencyanol)
in deionisiertem Formamid

II. METHODEN

1. Isolierung von RNA und Anreicherung von Poly(A)+RNA

Isolierung von RNA

RNA-Präparationen wurden mit Modifikationen nach der Methode von Adams et al. (1977) durchgeführt. Konfluente Zellen (ca. 5×10^6) in Petrischalen (9 cm Durchmesser) wurden durch eine 0.1 %ige Trypsin-Lösung abgelöst, wobei jeweils 10 Platten (von insgesamt 100 bis 150) gleichzeitig inkubiert wurden. Die Trypsinbehandlung wurde durch 5 % FKS (foetales Kalberserum) in PBS gestoppt, die Zellen wurden abzentrifugiert (7 min bei 1000 Upm) und mit PBS gewaschen. Zur Solubilisierung wurden pro g Zellen ca. 5 ml Guanidinium-Lösung (Puffer A) zugegeben. Zur Homogenisierung wurde die Suspension gepottet und durch Kanülen (mit abnehmendem Durchmesser) gepresst. Nach Entfernung von Zelldebris durch kurzes Abzentrifugieren (5 min bei 2000 Upm) wurde der Überstand mit einem halben Volumen Ethanol 30 min bei -20 °C gefällt. Anschließend wurde zentrifugiert (30 min bei 4500 Upm, 4 °C; HS-4 Rotor, Sorvall). Das Pellet wurde in Puffer B gelöst, und zwar in 1-2 ml pro g Zellen, und dann abermals wie oben gefällt und zentrifugiert. Dieser Vorgang wurde noch einmal wiederholt. Dann wurde das Pellet in 4 ml 20 mM EDTA gelöst und mit dem gleichen Volumen Chloroform/Isobutanol (4:1) extrahiert. Nach Reextraktion wurden zur wäßrigen Phase 3 Volumina 4 M Na-Acetat/pH 5-5.5 gegeben und über Nacht bei -15 °C präzipitiert. Die RNA wurde dann noch zweimal in 0.3 M Na-Acetat/pH 5-5.5 mit 2.5 Volumina Ethanol gefällt und anschließend in Wasser gelöst.

Zur Bestimmung der RNA-Konzentration wurde die Absorption bei 260 gemessen (1 OD = 40 µg RNA/ml). Außerdem wurde anhand der Absorptionen bei 260 nm und 280 nm der Reinheitsgrad der RNA festgestellt.

PBS: 140 mM NaCl

3 mM KCl

15 mM K-Phosphat

8 mM Na-Phosphat

pH 7.3

Puffer A: 8 M Guanidiniumhydrochlorid

2.5 mM Na-Acetat/pH 5-5.5

1 mM DTT

Puffer B: Puffer A + 20 mM EDTA

Anreicherung von Poly(A)+RNA an Oligo(dT)-Zellulose

Zur Anreicherung von Poly(A)+RNA wurden 3 bis 6 mg Gesamt-RNA (1mg/ml) über eine Säule mit Oligo(dT)-Zellulose (ca. 1 ml) geschickt, die mit Bindungspuffer äquilibriert war. Vor dem Auftrag wurde die RNA (in H₂O) 3-6 min bei 70 °C erhitzt und auf Eis abgekühlt. Dann wurde die RNA-Lösung an den Bindungspuffer angepaßt und auf die Säule aufgetragen. Nach drei Durchläufen wurde die Säule mit Bindungspuffer nachgewaschen und die gebundene RNA mit Elutionspuffer eluiert. Den Fraktionen (ca. 300 µl) des Eluats wurden 1-3 µl entnommen, um damit das Elutionsprofil entweder durch OD-Messung bei 260 nm oder durch Ethidiumbromidfärbung auf Tüpfelplatten (20 ml Agarose mit 10 µg Ethidiumbromid) festzustellen. Die gewonnene Poly(A)+RNA wurde noch ein zweites Mal über eine kleinere Oligo(dT)-Säule (ca. 0.5 ml) gegeben und nach drei Durchläufen wie oben eluiert. Ausbeute: 5-10 % der eingesetzten Gesamt-RNA.

Bindungspuffer: 20 mM Tris-HCl/pH 7.5

0.5 M NaCl

1 mM EDTA

Elutionspuffer: 10 mM Tris-HCl/pH 7.5

1 mM EDTA

2. Isolierung und Reinigung von DNA

Plasmidpräparation:

Plasmid-DNA wurde mit Modifikationen nach der Methode von Birnboim und Doly (1979) gewonnen.

Von 10-20 ml Übernachtskulturen wurden jeweils 2 ml in einem Eppendorfgefäß 1 min lang abzentrifugiert, das Pellet in 120 µl Lysispuffer gelöst und mit 80 µl Lysozymlösung (10 mg/ml Lysispuffer) versetzt. Nach 30 min Inkubation auf Eis wurden 400 µl einer frisch zubereiteten Lösung mit 0.2 N NaOH/1 % SDS dazugegeben, vorsichtig gemischt (durch mehrmaliges Umdrehen des Gefäßes) und die Ansätze 5 min auf Eis gestellt. Nach Zugabe von 300 µl 5 M K-Acetat/pH 4.8 wurde wiederum leicht gemischt und 30 min in Eiswasser inkubiert. Die präzipitierte chromosomale DNA wurde 15 min bei 4 °C abzentrifugiert (Eppendorfsentrifuge). Der Überstand wurde abgenommen und mit 60 % Isopropanol gefällt (30 min bei -20 °C). Nach Zentrifugation (15 min bei 4 °C) wurde das Pellet mit

80 % Ethanol gewaschen und in 100 µl 0.15 M NaCl, 0.05 M Tris-HCl/pH 8 gelöst. Der Inhalt mehrerer Eppendorfgläser wurde vereinigt und zur Entfernung von Protein mit Phenol/Chloroform (1:1, äquilibriert mit 0.15 M NaCl, 0.05 M Tris-HCl/pH 8) extrahiert. Die wässrige Phase wurde mit 2 Volumina Ethanol versetzt, 30 min bei -20 °C gefällt und abzentrifugiert (15 min bei 4 °C). Das Pellet wurde in 100 µl TE gelöst.

Zur RNA-Verdau wurde 0.1 µg RNase A zugegeben und 30 min bei 37 °C inkubiert. Nach erneuter Extraktion mit Phenol/Chloroform äquilibriert mit TE¹ wurde die Plasmid-DNA mit Ethanol gefällt und in TE gelöst (ca. 1µg DNA/µl).

Ausbeute: 50-150 µg

Lysispuffer: 25 mM Tris-HCl/pH 8

50 mM Glukose

10 mM EDTA

Teilweise wurde auch die schnellere Methode von Wilming (1985) angewandt, vor allem, wenn eine große Anzahl von Klonen zu analysieren war. Das Zellpellet einer 2 ml Übernachtskultur wurde in 200 µl TELT-Puffer suspendiert. Nach Zugabe von 20 µl einer frischen Lysosym-Lösung (10 mg/ml H₂O) wurde die Suspension 1 min in kochendem Wasser inkubiert und dann 5 min auf Eis abgekühlt. Anschließend wurde 8-10 min bei Raumtemperatur zentrifugiert ('Eppendorfzentrifuge') und der Überstand mit 2 Volumina Ethanol gefällt. Das Präzipitat wurde in 50 µl TE gelöst und zur Entfernung von Protein mit dem gleichen Volumen Phenol/Chlorophorm (1:1, äquilibriert mit TE) extrahiert. Die DNA in der wässrigen Phase wurde abermals gefällt und in 20 µl TE gelöst. Ausbeute: ca. 5 µg

TELT: 50 mM Tris-HCl/pH 7.5

62.5 mM EDTA

0.4 % Triton X-100

2.5 M LiCl

Präparation einzelsträngiger M13-DNA.

2 ml IS-Medium wurden mit 20 µl einer exponentiell wachsenden Bakterienkultur (IS+oli Stamm JM101 oder JM109) angeimpft und mit einem M13-Klon (Plaques von der Platte) inokuliert. Nach 8-10 h Schütteln bei 37 °C wurden die Bakterien abzentrifugiert (2 min Eppendorfzentrifuge). 1.6 ml des Phagehaltigen Überstandes wurden mit 3.2 ml PEG-Lösung (20 % PEG 6000, 3 M NaCl) versetzt und 15 min bei Raumtemperatur gefällt. Nach 10 min Zentrifugation ('Eppendorfzentrifuge') wurde der Überstand verworfen, die Gefäßwände ausgewischt und

Flüssigkeitsreste mit einer ausgezogenen Pasteurpipette abgenommen. Das gut getrocknete Pellet wurde dann in 100 µl TES-Puffer gelöst und mit dem gleichen Volumen Phenol/Chloroform (1:1) extrahiert. Die wässrige Phase wurde noch mehrmals (2-3x) mit Chloroform ausgeschüttelt, um restliches PEG zu entfernen. Dann wurde die DNA mit Ethanol gefällt und in 10 µl 10 mM Tris-HCl/pH 7.5 aufgenommen. Ausbeute: 2-5 µg

TES: 20 mM Tris-HCl/pH 7.5

10 mM NaCl

0.1 mM EDTA

Reinigung von DNA:

a) Säulenchromatographie

Vor einer geplanten Endmarkierung von DNA durch Polynukleotidkinase mußte niedermolekulare RNA möglichst quantitativ abgetrennt werden. Ca. 10 µg gespaltene Plasmid-DNA wurde über eine Säule (Durchmesser; 0.7 cm, Höhe: 20 cm) gegeben, die mit Sephacryl S-300 oder S-400 (in TE) gefüllt war. Von den gesammelten Fraktionen (150-200 µl) des Eluats wurden jeweils 2 µl abgenommen und der Plasmid-Gehalt in Miniagarosegelen oder auf Tüpfelplatten festgestellt.

b) Säulenzentrifugation

In Anlehnung an die von Maniatis et al. (1982) beschriebene Methode wurden Säulenchromatographien in 1 ml Plastikspritzen durchgeführt. Vor dem Füllen der Spritzen wurden die Ausflußöffnungen über einer Flamme verengt und mit silanisierter Glaswolle verstopft. Als Füllmaterial wurde Sephadex G-50 (Medium) in H₂O und Sephacryl S-300 in TE verwendet. Mit beiden Sorten wurden Salze und nicht eingebaute Nukleotide von DNA abgetrennt. S-300 wurde außerdem verwendet, um Polylinkerfragmente und niedermolekulare RNA abzutrennen. Vor der Applikation der Proben muß das Säulenmaterial gepackt werden, und zwar wurden S-300-Säulen 2 min bei 1000 Upm und G-50-Säulen 5 min bei 2000 Upm (Minifuge 2, Heraeus Christ) zentrifugiert. Unter den gleichen Bedingungen wurde dann die DNA (in 80-100 µl) zentrifugiert, wobei das Eluat in einem 1.4 ml-Eppendorffgefäß, das in ein Reagenzglas gesteckt war, aufgefangen wurde.

3. Restriktionen

Für DNA-Spaltungen wurden ausschließlich die von den Herstellerfirmen empfohlenen Puffer verwendet. Bei analytischen Spaltungen wurden folgende

Mengen eingesetzt: für Silberfärbungen von Fragmenten in Polyacrylamidgelen wurden 0.2-0.3 µg DNA mit 1 u Enzym in einem Volumen von 10 µl verdaut und der gesamte Ansatz auf ein Gel aufgetragen. Für Ethidiumbromidfärbungen in Miniagarosegelen wurden 1-2 µg DNA mit 2-4 u Enzym in 10-20 µl geschnitten und davon 1-3 µl analysiert. Bei präparativen Ansätzen wurden 10-15 µg Plasmid-DNA verdaut, in einem Volumen von 40-100 µl und mit Enzymmengen von 20-40 u. Um die Spaltungen zu kontrollieren, wurde den Restriktionsansätzen nach ca. 1.5 h ein Aliquot (0.5-1 µg DNA) entnommen und auf einem Miniagarosegel analysiert. War die Restriktion vollständig (nach 1.5 bis 3h), wurde die Reaktion durch Zugabe von EDTA (10-20 mM) gestoppt.

4. Gelelektrophorese, Anfärbung und Isolierung von DNA-Fragmenten

Gelelektrophorese unter nicht denaturierenden Bedingungen

Zur Bestimmung der Fragmentgrößen wurden bei der Gelelektrophorese jeweils Längenmarker mitaufgetragen, und zwar Hind III-gespaltene λ-DNA und/oder Hae II-geschnittene φx-DNA. Nach Restriktionen wurden die DNA-Proben vor dem Auftrag 10 min bei 65 °C erhitzt.

a) Agarosegele

Elektrophoresen von DNA-Proben wurden in horizontalen 0.8-1 %igen Agarosegelen (13.5 cm x 14/20 cm) mit dem Laufpuffer 1xTBE bei 100-150 V durchgeführt. Zur Analyse von Restriktionsspaltungen wurden Miniagarosegele (6.5x4 cm) verwendet (Auftrennung der DNA bei 75 V).

b) Polyacrylamidgele

Als Laufpuffer wurde hier generell 0.5 x TBE verwendet. 5 %ige Polyacrylamidgele wurden für die Auftrennung von DNA-Fragmenten benutzt, 15 %ige für die Analyse von Oligonukleotiden. Für die Isolierung von DNA-Fragmenten wurden Gele (20x25 cm) mit einer Dicke von 1.2 mm oder 2.4 mm verwendet. Die Elektrophorese wurde bei 300 V durchgeführt. Unter den gleichen Bedingungen wurden DNA-Proben vor einer Silberfärbung aufgetrennt, jedoch bei einer Geldicke von nur 0.8 mm. Teilweise wurden auch kleine Polyacrylamidgele (10x10 cm) verwendet, z. B. um Restriktionen radioaktiv markierter Fragmente oder die radioaktive Markierung von Oligonukleotiden zu kontrollieren.

Färbemethoden:

a) Ethidiumbromidfärbung

Agarose- und Polyacrylamidgele wurden in einer Ethidiumbromidlösung (5 µg/ml) gefärbt (10-20 min). Der Lösung war außerdem Ammoniumacetat (mindestens 0.1 M) zugesetzt, was die Färbung intensiviert. Die Gele wurden dann unter UV-Licht (302 nm) fotografiert (Polaroidfilm Typ 667 und Kodak 22A Wratten Filter).

b) Silberfärbung

Die für die Silberfärbung (nach Merril et al., 1981) von DNA-Fragmenten in Polyacrylamidgelen benötigten Lösungen wurden von der Firma Bio-Rad bezogen. Die Durchführung der Färbung erfolgte nach dem von der Herstellerfirma angegebenen Protokoll. Anschließend wurden die Gele auf einem Gelrockner auf Filterpapier (Schleicher & Schuell) getrocknet.

Isolierung von DNA-Fragmenten aus Polyacrylamidgelen:

DNA-Fragmente wurden in Polyacrylamidgelen aufgetrennt und durch Ethidiumbromidfärbung sichtbar gemacht. Bei radioaktiv markierten Fragmenten wurde die Position der Banden durch kurze Exponierung des Gels (10-30 min) mit einem Röntgenfilm festgestellt. Die DNA-Banden wurden aus den Gelen ausgeschnitten, und die Gelstücke in einen Dialyseschlauch gesteckt. In einer Agarosegelkammer wurden die Banden in 0.5 x TBE bei 150 V elektroeluiert (1-2 h). Anschließend wurden die Elektroden kurz (30 sec) umgepolt, um die DNA von der Membran zu lösen. Die Lösung (ca. 0.5 ml) wurde entnommen und mit Isobutanol auf 100 µl eingeeengt. Durch Säulenzentrifugation wurden die Proben von Salz, Arcylamid- und Butanolresten befreit.

5. RNA-Gele und Northernblot

Methylquecksilberhydroxid-Gele

3 µl RNA-Lösung (2 µg/µl) wurden mit 3 µl 2xBS-Puffer mit 20 mM Methylquecksilberhydroxid, sowie 10 % Glycerin und 0.05 % Farbstoff (Bromphenolblau und Xylencyanol) 1 min bei 50 °C denaturiert und auf das Gel aufgetragen. Das Gel (0.8-1 %ige Agarose, 13x14 cm) enthielt ca. 6 mM Methylquecksilberhydroxid. Die Elektrophorese wurde 3-4 h bei 80-100 V in 1x BS-Puffer durchgeführt. Anschließend wurde das Gel zur Entfernung von Methylquecksilberhydroxid 3x 20 min in 0.4 M Ammoniumacetat geschüttelt und 2x 15 min in 0.5xSSC gewaschen. Die RNA wurde durch Kapillartransfer auf Filter (Genescreen, NEN) übertragen wie von Maniatis et al. (1982) beschrieben, mit dem Unterschied, daß das Pufferreservoir

mit 20xSSC gefüllt wurde. Nach dem RNA-Transfer (20-24 h) wurden die Filter an Luft getrocknet und 2-3 h bei 80 °C gebacken.

Alternativ wurden auch Agarosegele nach der Anleitung von Tsao et al. (1983) direkt für die Hybridisierung eingesetzt. Hierbei wurde das Gel nach der Elektrophorese wie oben behandelt und dann auf einem Gelrockner auf eine Dialysemembran aufgetrocknet. Zur Ablösung des Gels wurde die Membran kurz in Wasser eingetaucht.

BS: 50 mM Borsäure
5 mM Na-Tetraborat
10 mM Na-Sulfat

Glyoxalgele

Ca. 10 µg RNA wurden mit Ethanol gefällt und in 8 µl Glyoxal-Puffer gelöst. Die RNA wurde 1 h bei 50 °C im Wasserbad denaturiert und auf Eis abgekühlt. Nach Zugabe von Auftragspuffer (10 % Glycerin, 2 mM Natriumphosphat/pH 6.8 und 0.05 % Farbstoff) wurde die RNA in einem 0.8-1 %igen Agarosegel (13x20 cm) bei 90 V aufgetrennt. Der Laufpuffer (10 mM Natriumphosphat/pH 7) wurde während der 6 h dauernden Elektrophorese durch eine Pumpe umgewälzt.

Anschließend wurde die RNA durch Elektroblotting auf Filter (Genescreen) übertragen, und zwar ca. 16 h lang bei 0.7 mA, 4 °C, in 10 mM Na-Phosphat/pH 7. Nach Trocknung an Luft wurde der Filter dann 2-3 h bei 80 °C gebacken. Vor der Hybridisierung wurde der Filter zur Entfernung von Glyoxal in einem Plastikbeutel 5-10 min bei 100 °C erhitzt und auf Raumtemperatur abgekühlt.

Glyoxalpuffer: 1 M Glyoxal, deionisiert
50 % DMSO
10 mM Na-Phosphat, pH 7

6. Southernblot

DNA wurde in 0.8 %igen Agarosegelen in 1xTBE aufgetrennt. Die Gele wurden dann 20 min in 0.5 M NaOH/0.15 NaCl und 20 min in 0.5 M Tris-HCl/pH 8, 0.15 M NaCl inkubiert und anschließend in 0,5xSSC gewaschen. Die DNA wurde entweder durch Kapillartransfer auf Filter übertragen (Maniatis et al., 1982), oder das Gel wurde getrocknet und direkt für die Hybridisierung eingesetzt (siehe 5.).

7. Radioaktive Markierung von DNA

Markierung der 5'-Enden mit Polynukleotidkinase:

a) DNA-Fragmente

Zur Entfernung des 5'-Phosphats wurden isolierte DNA-Fragmente (2-5 μg) in einem Volumen von 20 μl (0.1 M Tris-HCl/pH 7.9) mit 0.5 u CIP ("Calf intestine alkaline phosphatase") inkubiert (20 min bei 37 °C, dann 10 min bei 45 °C). Anschließend wurden 2.2 μl 0.1 M Nitrilotriessigsäure dazupipettiert und die Lösung 20 min bei Raumtemperatur stehengelassen. Nach Zugabe von 1.6 μl Spermidin (5 mM) wurde die Probe für 2 min bei 95 °C erhitzt. Durch diese Prozedur wird die CIP irreversibel inaktiviert. Das Enzym muß daher nicht entfernt werden, so daß die nachfolgende Kinase-Reaktion im gleichen Ansatz stattfinden kann.

Nach Zugabe von 2 μl 20x Puffer, 5-10 μl ^{32}P - γ -ATP (10 $\mu\text{Ci}/\mu\text{l}$, 3000 Ci/mmol) und 20 u Polynukleotidkinase wurde in einem Volumen von 40 μl 30 min bei 37 °C inkubiert. Die Kinasereaktion wurde durch 1 μl 0.5 M EDTA gestoppt. Nach Extraktion mit dem gleichen Volumen Phenol/Chloroform (1:1) wurde die wässrige Phase über eine Sephadex G-50-Säule zentrifugiert, um freies ^{32}P - γ ATP abzutrennen. Ausbeute: $1\text{-}5 \times 10^6$ cpm

20x Kinasepuffer: 0.1 M Tris-HCl/pH 7.7

0.5 M MgCl_2

0.5 M β -Mercaptoäthanol

b) Oligonukleotide

Oligonukleotide werden ohne 5'-Phosphat synthetisiert, daher entfällt die Phosphatasebehandlung.

10-15 pmol der Oligonukleotide wurden mit 7-10 μl ^{32}P - γ -ATP (10 $\mu\text{Ci}/\mu\text{l}$, 3000 Ci/mmol) und 4 μl 5x Puffer in einem Endvolumen von 20 μl durch 10 u Polynukleotidkinase markiert. Nach 30 min Inkubation bei 37 °C wurde die Reaktion mit 1 μl 0.5 M EDTA gestoppt und die Probe mit Phenol/Chloroform (1:1) extrahiert. Nicht eingebaute Nukleotide wurden wiederum durch Säulenzentrifugation (Sephadex G-50) entfernt. Da die verwendeten 20 Basen langen Oligonukleotide bei nur einer Zentrifugation nicht vollständig aus den Sephadex G-50-Säulen eluiert wurden, wurden nach dem ersten Lauf nochmals 100 μl H_2O auf die Säulen aufgetragen und ein zweites Mal zentrifugiert. Ausbeute: $10\text{-}50 \times 10^6$ cpm

5x Kinasepuffer: 0.5 M Tris-HCl/pH 7.7

0.1 M MgCl_2

1 mM EDTA

50 mM DTT

1 mM Spermidin

Nicktranslation:

Die Ansätze (20 µl) enthielten 0.1-0.2 µg der zu markierenden DNA, 2 µl 10x Nick-translationspuffer, je 30 µM der unmarkierten dNTPs, 7 µM ATP, sowie 100 µCi 32 P- α -dNTP. Zunächst wurden 200 pg DNase zugegeben und 5 min bei 14 °C inkubiert. Dann wurden 10 u DNA-Polymerase I dazupipettiert und die Reaktion 60-90 min bei 14 °C durchgeführt. Anschließend wurde mit 1 µl 0.5 M EDTA gestoppt, mit Phenol/Chloroform (1:1) extrahiert und der Ansatz über eine Sephadex G-50-Säule zentrifugiert. Vor Zugabe zur Hybridisierungslösung wurden die nicktranslatierten Proben 5 min bei 95 °C erhitzt. Ausbeute: $1-10 \times 10^6$ cpm

10x Nicktranslationspuffer: 0.5 M Tris-HCl/pH 7.8

0.1 M $MgCl_2$

10 mM DTT

Markierung von einzelsträngiger M13-DNA:

2 µl (0.5-1 µg) einzelsträngige M13-DNA wurden mit 2 µl (3 ng) Hybridisierungsprimer, 1 µl 10x Puffer und 3 µl H_2O 10 min bei 55 °C erhitzt. Dann wurde der Ansatz langsam (30 min) auf Raumtemperatur abgekühlt. Es wurden 20 µCi 32 P- α -dNTP (10 µCi/µl, 3000 Ci/mmol), sowie je 2 µl der anderen dNTPs (0.5 M), 1 µl 10x Puffer und 1 u DNA-Polymerase (Klenow-Fragment) zugegeben. In einem Volumen von 20 µl wurde 60 min bei 15 °C inkubiert. Nach Beendigung der Reaktion durch EDTA wurde mit Phenol/Chloroform extrahiert und eine Säulenzentrifugation (Sephadex G-50) angeschlossen. Ausbeute: $2-10 \times 10^6$ cpm

10x Puffer: 60 mM Tris-HCl/pH 7.5

60 mM NaCl

60 mM $MgCl_2$

10 mM DTT

8. Hybridisierungen

Einsatz von Oligonukleotiden

Northernhybridisierungen mit Oligonukleotiden erfolgten nach Breslow et al. (1982) in 0.15 M Tris-HCl/pH 8, 0.75 M NaCl, 10 mM EDTA, 5x Denhardt's-Lösung, 0.1 % SDS und 100 µg/ml Heringssperma-DNA, Southernhybridisierungen mit Oligonukleotiden in 5xSSC, 10x Denhardt's-Lösung, 0.1 % SDS und 10 µg/ml Heringssperma-DNA. 100x Denhardt's-Lösung enthält jeweils 2 % Ficoll, Polyvinylpyrrolidon und BSA. Die Wahl eines geeigneten Temperaturbereichs für die Hybridisierung und das Waschen der Filter wurde in Teil I,2. (S. 18) eingehend be-

schrieben. Die Filter oder Gele wurden zunächst 6 h bei der gewünschten Temperatur (40-53 °C) in Plastikbeuteln prähybridisiert (in 3-5 ml). Dann wurden die radioaktiv markierten Oligonukleotide ($1-20 \times 10^6$ cpm) zugegeben, der gesamte Ansatz 10 min auf 65 °C erhitzt und wieder auf die Hybridisierungstemperatur abgekühlt. Nach der Inkubation (mindestens 16 h) wurden die Filter (Gele) 3x 30 min in 6xSSC/1 % SDS (Northern) oder 5xSSC/ 0.1 % SDS (Southern) gewaschen, und zwar bei Hybridisierungstemperatur und höheren Temperaturen (bis 53 °C). Die noch feuchten Filter wurden in einen Plastikbeutel eingeschweißt und mit einem Röntgenfilm und einer Verstärkerfolie bei -70 °C ein bis mehrere Tage exponiert.

Einsatz von nicktranslatierten Fragmenten und markierter M13-DNA

Sowohl Northern- wie auch Southernhybridisierungen erfolgten in 5xSSC mit 1 % Sarkosyl. Nach der Prähybridisierung (1 h) wurde die Lösung ausgetauscht und die Probe ($1-5 \times 10^6$ cpm) zugegeben. Nach der Inkubation (16-24 h) wurde 2x 30 min mit 3xSSC/0.5 % Sarkosyl und 2x 30 min mit 3xSSC gewaschen.

9. "screening" von cDNA-Genbanken

60 000 bis 100 000 Transformanten einer cDNA-Genbank wurden auf 30-40 Nitrozellulosefilter (Durchmesser 82 mm) ausplattiert und über Nacht angezüchtet. Für die EHS-Genbank (Vektor: pBR322) wurden Platten mit Tetracyclin (15 µg/ml) verwendet, für die PYS-2-Genbanken (pUC-Vektoren) Platten mit Ampicillin (100 µg/ml). Von jedem Filter wurde ein Replika hergestellt, wobei beide Filter durch Nadeleinstiche am Rand markiert wurden. Bei Arbeiten mit der EHS-Genbank wurden die Replika-Filter nach dem Hochwachsen der Kolonien zur Plasmid-Amplifizierung auf Platten mit Chloramphenicol (150 µg/ml) transferiert und ca. 40 h inkubiert.

Zur Lysis der Bakterien wurden die Filter mit folgenden Lösungen behandelt (durch Auflegen auf Filterpapier, das mit der jeweiligen Lösung angefeuchtet war): 3 min auf 10 % SDS, 10 min auf 0.5 N NaOH, 10 min auf 1 M Tris-HCl/pH 8, 0.6 M NaCl, 30 min auf 1 M Tris-HCl/pH 8, 0.6 M NaCl. Beim letzten Schritt wurden unter jeden Filter 20 µl einer Proteinase K-Lösung (20 mg/ml H₂O) pipettiert. Zur Entfernung von Protein und Zellresten wurden die Filter mit Chloroform behandelt (30 sec auf Chloroform schwimmen lassen, mit der Kolonien-seite nach oben) und dann mit einer Flasche unter starkem Druck zwischen zwei Rundfiltern (Schleicher & Schuell) abgerollt. Nach Trocknung an Luft wurden die Nitrozellulosefilter unter Vakuum 2-3 h bei 80 °C gebacken.

Die Hybridisierungen und das Waschen der Filter erfolgte dann wie unter 8. beschrieben. Alle Filter wurden übereinander gestapelt in einem einzigen Plastikbeutel in 30-40 ml Lösung inkubiert. Nach der Prähybridisierung wurde generell frische Lösung zugesetzt. Bei Hybridisierungen mit synthetischen Oligonukleotiden wurden $10\text{-}20 \times 10^6$ cpm zugegeben, beim "screening" mit nicktranslatierter DNA $1\text{-}10 \times 10^6$ cpm.

Nach Exponierung der Filter (1-5 Tage je nach Intensität der Signale) wurden für positive Signale anhand der Einstich-Markierungen auf den Filtern die entsprechenden Kolonien auf den Ausgangsplatten gesucht und mit einem Zahnstocher auf eine neue Platte gepickt (siehe auch Abb.23, S. 92). Wenn es wegen zu großer Zelldichte nicht möglich war, Einzelkolonien zu identifizieren und zu treffen, wurden ca. 10 Kolonien aus dem betreffenden Areal gepickt. Replikas der Filter mit den gepickten Klonen wurden dann wie oben weiterverarbeitet und hybridisiert. Häufig handelte es sich bei den gepickten Kolonien um Mischkolonien, so daß eine zweite Nachreinigung notwendig war. Hierzu wurden die fraglichen Kolonien in Medium soweit verdünnt, daß nach dem Ausplattieren eines Aliquots Einzelkolonien (50-200 pro Platte) vorhanden waren. Die positiven Kolonien wurden durch erneute Hybridisierung identifiziert.

10. Klonierung von DNA

Ligationen:

Ca. 1 µg des jeweiligen Vektors (pUC9, M13tg130, M13tg131) wurden mit Restriktionsenzymen verdaut. Bei Einfachspaltungen wurde im Anschluß an die Restriktion alkalische Phosphatase zugesetzt (0.5 u CIP). Nach 30 min Inkubation wurde die Phosphatase wie unter 7. beschrieben durch Nitrilotriessigsäure und Erhitzen inaktiviert. Der Ansatz wurde mit Phenol/Chloroform extrahiert und über eine Sephadex G-50-Säule gereinigt. Nach Doppelspaltungen wurden die Ansätze über eine Sephacryl S-300-Säule zentrifugiert, um die entstandenen kurzen Polylinkerfragmente abzutrennen.

Für Klonierungen in das Plasmid pUC9 wurden für die Ligationen 0.05-0.1 µg Vektor und 0.02-0.05 µg Fragment-DNA eingesetzt, für Klonierungen im M13-Vektor ca. 0.1 µg M13-DNA (replikative Form) und 0.05-0.1 µg Fragment-DNA. In einem Volumen von 20 µl wurde nach Zugabe von 1 u Ligase 1 h bei 4 °C und dann über Nacht bei 16 °C inkubiert.

10x Ligationspuffer: 0.5 M Tris-HCl/pH 7.4
0.1 M $MgCl_2$
0.1 M DTT
10 mM Spermidin
10 mM ATP

Transformation von E.coli:

Für Plasmidtransformationen wurde die Stämme JM83, DH1 und DH5 verwendet, für Transformationen mit M13-Vektoren die Stämme DH1 und DH5. DH1 und DH5 haben den Vorteil, daß sehr hohe Transformationsraten erreicht werden können (bis zu 5×10^8 /µg DNA, nach Hanahan, 1983). Allerdings exprimieren diese Stämme eine funktionelle β -Galaktosidase, so daß positive Klone nicht anhand des Farbttests mit X-Gal identifiziert werden können. Außerdem besitzen sie keine F-Pili und können daher nicht von M13-Phagen infiziert werden. Nach Transformation der DH1- oder DH5-Zellen mit rekombinanter M13-DNA mußten daher Indikatorzellen (JM101, JM109) zugegeben werden.

a) Herstellung kompetenter Zellen

Eine Übernachtskultur des Stammes JM83 wurde 1:100 in LB-Medium verdünnt und bis zu einer Zellzahl von 5×10^7 /ml angezüchtet (Die entsprechenden OD-Werte wurden einer Wachstumskurve entnommen, die für jeden einzelnen Stamm bestimmt worden war). Die Bakteriensuspension wurde 10 min auf Eis abgekühlt und die Zellen abzentrifugiert. Das Pellet wurde in 50 mM $CaCl_2$, 10 mM Tris-HCl/pH 7.9 resuspendiert (in einem Drittel des Ausgangsvolumens). Nach erneuter Zentrifugation wurden die Zellen in einem Dreißigstel des gleichen Puffers suspendiert und 20-24 h auf Eis gesetzt.

Transformationsrate: $4 \times 10^6 - 10^7$ /µg DNA

Kompetente Zellen der Stämme DH1 und DH5 wurden nach der Methode von Hanahan (1985) hergestellt. 100-400 ml TYM-Medium wurden direkt von einer Platte angeimpft und bis zu einer Zellzahl von 5×10^7 /ml hochgezüchtet. Die Zellen wurden bei 4 °C abzentrifugiert und in 30 ml/100 ml Ausgangsvolumen TfbI-Puffer zur schnellen Resuspendierung gevortext. Die Zellen wurden erneut abzentrifugiert und in 4 ml/100 ml Ausgangsvolumen TfbII-Puffer durch leichtes Schütteln auf Eis suspendiert. Die kompetenten Zellen wurden aliquotiert und bei -70 °C eingefroren. Transformationsrate: $5 \times 10^7 - 1 \times 10^8$

Tfbl: 30 mM Ka-Acetat

50 mM MnCl₂

100 mM KCl

10 mM CaCl₂

15 % Glycerin

pH 5.8

TfbII: 10 mM Na-MOPS/pH 7

75 mM CaCl₂

10 mM KCl

15 % Glycerin

b) Transformation mit pUC-Plasmiden

200 µl kompetente JM83-Zellen wurden auf Eis mit 10 µl eines Ligationsansatzes versetzt, bzw. 20 µl kompetente DH1- oder DH5-Zellen mit 1 µl eines Ligationsansatzes. Nach 20 min auf Eis wurden die Ansätze 90 sec bei 37 °C erhitzt und 1 min auf Eis abgekühlt. Dann wurden 0.5 ml LB- oder 2xYT-Medium zugegeben und zur Ausbildung der Ampicillinresistenz 50 min bei 37 °C inkubiert. 200-400 µl der Suspension wurden auf LB-platten mit Ampicillin (100µg/ml) und X-Gal (40 µg/ml) ausplattiert.

c) Transformation mit rekombinanter M13-DNA

20 µl kompetenter Zellen (DH1,DH5) wurden mit 1 µl eines Ligationsansatzes versetzt und 30 min in Eis gestellt. Nach einem Hitzeschock (90 sec bei 37 °C) wurden die Ansätze wieder auf Eis abgekühlt. Dann wurden 100 µl exponentiell wachsender Indikatorzellen (JM101, JM109), sowie 3 ml Weichagar (45 °C) mit 10 µl 10 mM IPTG und 25 µl 4 % X-Gal zugegeben und auf einer YT-Platte verteilt. Nach 8-10 h bei 37 °C waren Plaques mit rekombinanten Phagen anhand fehlender Blaufärbung zu erkennen.

11. Sequenzierung von DNA

Chemische Methode nach Maxam und Gilbert (1980)

Die Reaktionen und weitere Aufarbeitung der DNA erfolgte im Wesentlichen nach dem Protokoll von Maxam und Gilbert (1980).

Die ³²P-endmarkierte DNA (5x10⁴ bis 1x10⁶ cpm) wurde zusammen mit 5 µg Heringssperma-DNA mit Ethanol gefällt und in 32 µl H₂O gelöst. Davon wurden je 6 µl auf fünf silanisierte Eppendorfgefäße verteilt. Die Reaktionen wurden nach folgendem Schema durchgeführt:

Basenspezifische Reaktionen	G	G+A	C+T	C	A>C
Ansätze: DNA-Probe	6 µl	6 µl	6 µl	6 µl	6 µl
H ₂ O	-	-	20 µl	-	-
	200 µl	-	-	20 µl 5 M NaCl	-
DMS-Puffer					
Reagenz	1 µl DMS	25 µl Ameisensäure	20 µl Hydrazin	20 µl Hydrazin	100 µl 1,2 N NaOH/1 mM EDTA
Inkubation:	5 min bei 14 °C	jeweils 6 min bei 20 °C			6 min bei 95 °C
Stop	50 µl DMS- Stopplösung	je 200 µl Hydrazin-Stopplösung			150 µl 1 N Essig- säure
Zugabe von Ethanol	750 µl	750 µl	630 µl	630 µl	750 µl

Nach der Zugabe von Ethanol wurden die Ansätze 5-10 min in Trockeneis gestellt und 10 min bei 4 °C zentrifugiert. Die Überstände wurden abgenommen und Flüssigkeitsreste mit einer Glaskapillare entfernt. Die Pellets wurden in 200 µl 0.2 M Na-Acetat/pH 5-5.5 gelöst und mit 500 µl Ethanol gefällt. Nach Zentrifugation wurde dieser Schritt wiederholt. Die getrockneten Pellets wurden dann in 25 µl 1 M Piperidin gelöst. Nach der Piperidinspaltung (30 min bei 90 °C) wurde Piperidin nicht durch Lyophilisierung entfernt, wie im ursprünglichen Protokoll angegeben, sondern durch zweimalige Ethanolfällung. Zuerst wurden 50 µl 0.4 M Na-Acetat/pH 5-5.5 zugesetzt und mit dem 2.5-fachen Volumen Ethanol gefällt. Die DNA wurde dann in 100 µl 0.2 M Na-Acetat/pH 5-5.5 aufgenommen und ein zweites Mal mit Ethanol gefällt. Danach wurde das getrocknete DNA-Pellet in 5 µl Auftragspuffer gelöst und davon 1-2 µl auf ein Sequenzgel aufgetragen.

DMS-Puffer: 50 mM Na-Cacodylat/pH 8
1 mM EDTA

DMS-Stopplösung: 1.5 M Na-Acetat/pH 7
1 M Mercaptoethanol
100 µg/ml tRNA

Hydrazin-Stopplösung: 0.3 M Na-Acetat/pH 7
0.1 mM EDTA
25 µg/ml tRNA

Dideoxy-Methode nach Sanger (1980)

3 µl einzelsträgige M13-DNA (0.5-1 µg) wurden mit 2µl (3 ng) Sequenzierungsprimer und 1 µl 10x Puffer (100 mM Tris-HCl/pH 7.5, 100 mM NaCl) 10 min bei 55 °C erhitzt und langsam (30 min) auf Raumtemperatur abgekühlt. Nach Zugabe von 2-4 µl ³²S-α-ATP (10 µCi/µl, 500 Ci/mMol) wurden 2 µl auf vier Ansätze mit jeweils 3 µl der basenspezifischen Reaktionsmischungen verteilt. Diese Mischungen (2x Lösungen) enthielten neben 10 mM NaCl, 10 mM MgCl₂, 10 mM Tris-HCl/pH 7.5 und 1 mM DTT die in der Tabelle angegebenen Nukleotidkonzentrationen (Diese Nukleotidkonzentrationen mußten an die Basenzusammensetzung des jeweiligen Templates angepaßt werden). Dann wurde jeweils 1 u DNA-Polymerase (Klenow-Fragment) an die Gefäßwände pipettiert und die vier Reaktionen durch kurzes Abzentrifugieren gleichzeitig gestartet. Nach 20 min Inkubation bei Raumtemperatur wurde 1 µl "Chase"-Lösung (je 2 mM dNTPs, 10 mM Tris-HCl/pH 7.5) zugesetzt und weitere 15 min inkubiert. Die Reaktionen wurden durch 3 µl Auftragspuffer gestoppt.

Basenspezifische Mischungen

Sequenzierung	M13				pUC			
(µM)	G	A	T	C	G	A	T	C
ddGTP	3,3-16,6	-	-	-	65-100	-	-	-
ddATP	-	70-140	-	-	-	50-100	-	-
ddTTP	-	-	117	-	-	-	500	-
ddCTP	-	-	-	16,6-33	-	-	-	85
dGTP	1,66	33	33	33	5	100	100	100
dATP	-	-	-	-	-	-	-	-
dTTP	33	33	1,66	33	100	100	5	100
dCTP	33	33	33	1,66	100	100	100	10

Bei der Doppelstrang-Sequenzierung nach Chen und Seeburg (1985) wurden 2 µg Plasmid-DNA in 32 µl H₂O mit 4 µl 2 N NaOH und 4 µl 2 mM EDTA versetzt und 5 min bei Raumtemperatur denaturiert. Nach Zugabe von 4 µl 2 M Ammoniumacetat/pH 4.5 und 88 µl Ethanol wurde die DNA gefällt (15 min bei -70 °C). Dann wurde 15 min zentrifugiert (4 °C, Eppendorfizentrifuge) und mit 80 % Ethanol gewaschen. Zu dem getrockneten Pellet wurden 1.5 µl 10x Puffer (70 mM Tris-HCl/pH 5.5, 70 mM MgCl₂, 50 mM Mercaptoethanol, 1 mM EDTA), 2 µl Sequenzierungsprimer (3 ng), 2 µl ³²S-α-ATP und 8.5 µl H₂O pipettiert. Der Ansatz wurde 15 min bei 37 °C inkubiert. Anschließend wurden jeweils 3 µl auf vier

Eppendorfgeläße mit je 2 µl der basenspezifischen Reaktionsmischungen verteilt. Diese Mischungen (3x Lösungen) enthielten die gleichen Pufferkomponenten wie die M13-Mischungen (in dreifacher Konzentration) und die in der Tabelle aufgeführten Nukleotidkonzentrationen. Nach Zugabe von je 1 u DNA-Polymerase fand die Inkubation unter den gleichen Bedingungen wie bei der M13-Sequenzierung statt.

Sequenzgele

Nach der Methode von Ansorge und Barker (1984) wurden die Sequenzgele kovalent an eine Glasplatte gebunden und während der Elektrophorese thermostatisiert (LKB-Makromould-Apparatur):

Die Thermostatisierplatte wurde mit 2x 5 ml "Repel"-Silan (2 % Dichlordimethylsilan in Trichloräthan) vorbehandelt, die Gelplatte mit "Bind"-Silan (0.3 % γ-Methacryloxypropyltrimethoxysilan in 15 ml Ethanol/0.3 % Essigsäure). Die Gellösung (50 ml) enthielt 21 g Harnstoff (7 M), 2.38 g Acrylamid, 0.12 g Bisacrylamid, 5 ml 10xTBE, 300 µl 10 % Ammoniumpersulfat und 20 µl TeMED. Als Laufpuffer wurde 1xTBE eingesetzt.

Während der Präelektrophorese (1500 V) wurde das Gel (Dicke: 0.2 mm) durch die Thermostatisierplatte auf 60-65 °C aufgeheizt. Die Sequenzierungsproben wurden 2 min bei 95 °C denaturiert und jeweils 1-2 µl mit einer Glaskapillare in die Geltaschen appliziert. Die Auftrennung der Proben erfolgte dann bei 3000 V. Die Elektrophorese wurde beendet, wenn der Farbstoff Bromphenolblau die Gelgrenze erreicht hatte (nach ca. 1 h). Mehrmalige (bis zu drei) Aufträge einer Probe erfolgten im Zeitabstand von 60-80 min. Nach der Elektrophorese wurde das auf der Gelplatte haftende Sequenzgel in 10 % Essigsäure fixiert (15 min), 90 min bei 80 °C getrocknet und mit einem Röntgenfilm in einer Kassette bei -70 °C exponiert (über Nacht).

12. Aufarbeitung von Oligonukleotiden

Die synthetischen Oligonukleotide Oligo I und Oligo II wurden dankenswerterweise von Dr. J. Engels (Hoechst AG) und Prof. Dr. D. Oesterheldt (Martinsried) synthetisiert.

Die Oligonukleotidsynthese erfolgte nach der Phosphoamidit-Methode (Beaucage und Caruthers, 1981) in einem 380 A DNA-Synthesizer (Applied Biosystems).

Die Syntheseprodukte enthielten noch die Basenschutzgruppen, sowie eine 5'-O-Dimethyltritylgruppe. Zur Abspaltung der Basenschutzgruppen wurden die Oligo-

nukleotide in konzentriertem Ammoniak inkubiert (5 h bei 56 °C). Zur Entfernung des Ammoniaks wurden die Proben mit 1 ml 0.1 M Ammoniumacetat/pH 5.4 versetzt und lyophilisiert. Anschließend wurden die Oligonukleotide in 0.2 ml 0.1 M Ammoniumacetat/pH 5.4 gelöst. Um kurzerkettige Syntheseprodukte abzutrennen und zur Reinigung der Oligonukleotide wurde eine "reversed phase" HPLC an einer VYDAC C18-Säule (LATEK, Heidelberg) mit einem Acetonitrilgradienten (5-60 % in 0.1 M Ammoniumacetat/pH 5.4) durchgeführt. Die Methode basiert darauf, daß nur die Endprodukte eine Tritylgruppe enthalten und daher von unvollständig synthetisierten kürzeren Oligonukleotiden (ohne Tritylgruppe) aufgrund veränderter Polarität abgetrennt werden können. Die "peak"-Fraktionen mit den Oligonukleotiden der richtigen Länge wurden wiederum lyophilisiert. Die Abspaltung der 5'-Tritylgruppe erfolgte in 1 ml 80 % Essigsäure 15 min bei Raumtemperatur. Nach dreimaligem Ausschütteln mit Ether wurde 1 ml 0.1 M Ammoniumacetat/pH 5.4 zugegeben und die Proben lyophilisiert. Die Oligonukleotide wurden dann dreimal mit jeweils 300 µl Methanol in einer Speed-Vac-Zentrifuge eingetrocknet und in 20-30 µl H₂O gelöst. Der Reinheitsgrad und die Konzentration der Oligonukleotide wurde durch ein UV-Spektrum (220-400 nm) festgestellt.

13. Computeranalyse

Zur Verarbeitung von DNA- und Proteinsequenzdaten wurden die Computerprogramme Naq, Relate und Align (George et al., 1982; Barker et al., 1983), sowie UWGGC-Programme (Devereux et al., 1984) verwendet.

G. LITERATUR

- Adams, S. L., Sobel, M. E., Howard, B. H., Olden, K., Yamada, K. M., de Crombrughe, B. and Pastan, I. (1977)
"Levels of translatable mRNAs for surface protein, collagen precursors and two membrane proteins are altered in Rous sarcoma virus-transformed chick embryo fibroblasts"
Proc. Natl. Acad. Sci. USA 74, 3399-3403
- Aho, S., Tate, V. and Boedtker, H. (1983)
"Multiple 3' ends of the chicken pro $\alpha 2(I)$ collagen gene"
Nucl. Acids Res. 11, 5443-5450
- Alitalo, K., Vaheri, A., Krieg, T. and Timpl, R. (1980)
"Biosynthesis of Two Subunits of Type IV Procollagen and of Other Basement Membrane Proteins by a Human Tumor Cell Line"
Eur. J. Biochem. 109, 247-255
- Amara, S. G., Evans, R. M. and Rosenfeld, M. G. (1984)
"Calcitonin/Calcitonin Gene-Related Peptide Transcription Unit: Tissue-Specific Expression Involves Selective Use of Alternative Polyadenylation Sites"
Mol. Cell. Biol. 4, 2151-2160
- Ansorge, W. and Barker, R. (1984)
"System for DNA sequencing with resolution of up to 600 base pairs"
J. Biochem. Biophys. Meth. 9, 33-48
- Appleyard, R. K. (1954)
"Segregation of new lysogenic types during growth of a doubly lysogenic strain derived from *Escherichia coli* K12"
Genetics 39, 440-452
- Babel, W. and Glanville, R. W. (1984)
"Structure of human-basement-membrane (type IV) collagen
Complete amino-acid sequence of a 914-residue-long pepsin fragment from the $\alpha 1(IV)$ chain"
Eur. J. Biochem. 143, 545-556
- Bächinger, H. P., Fessler, L. I. and Fessler, J. H. (1982)
"Mouse Procollagen IV. Characterization and Supramolecular Association"
J. Biol. Chem. 257, 9796-9803
- Barker, W. C., Hunt, L. T., Orcutt, B. C., George D. G., Veh, L. S., Chen, H. R., Blomquist, M. C., Johnson, G. C. and Dayhoff, M. D. (1983)
In Atlas of Protein Sequence and Structure. Protein Sequence Data Base, Version 7. National Biomedical Research Foundation. Washington, D. C.
- Barsh, G. S., Roush, C. L. and Gelinas, R. E. (1984)
"DNA and Chromatin Structure of the Human $\alpha 1(I)$ Collagen Gene"
J. Biol. Chem. 259, 14906-14913

- Beaucage, S. L. and Caruthers, M. H. (1981)
"Deoxynucleoside phosphoramidites: a new class of key intermediates for deoxypolynucleotide synthesis"
Tetrahedron Letters, 22 1859-1862
- Bennetzen, J. L. and Hall, B. D. (1982)
"Codon Selection in Yeast"
J. Biol. Chem. 257, 3026-3031
- Berg, R. A. and Prockop, D. J. (1973)
"The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple helix of collagen"
Biochim. Biophys. Res. Commun. 52, 115-120
- Berget, S. M. (1984)
"Are U4 small nuclear ribonucleoproteins involved in polyadenylation?"
Nature 309, 179-181
- Bilanchone, V., Duester, G., Edwards, Y. and Smith, M. (1986)
"Multiple mRNAs for human alcohol dehydrogenase (ADH): developmental and tissue specific differences"
Nucl. Acids Res. 14, 3911-3925
- Birnboim, H. C. and Doly, J. (1979)
"A rapid alkaline extraction procedure for screening recombinant plasmid DNA"
Nucl. Acids Res. 7, 1513-1523
- Birnstiel, M. L., Busslinger, M. and Strub, K. (1985)
"Transcription Termination and 3' Processing: The End Is in Site!"
Cell 41, 349-359
- Boardman, M., Basi, G. S. and Storti, R. V. (1985)
"Multiple polyadenylation sites in a *Drosophila* tropomyosin gene are used to generate functional mRNAs"
Nucl. Acids Res. 13, 1763-1776
- Boedtke, H., Fuller, F. and Tate, V. (1983)
"The Structure of Collagen Genes"
Int. Rev. Connect. Tissue Res. 10, 1-61
- Breslow, J. L., McPherson, J., Nussbaum, A. L., Williams, H. W., Lofquist-Kahl, F., Karathanasis, S. K. and Zannis, V. I. (1982)
"Identification and DNA Sequence of a Human Apolipoprotein E cDNA Clone"
J. Biol. Chem. 257, 14639-14641
- Brinker, J. M., Gudas, L. J., Loidl, H. R., Wang, S., Rosenblom, J., Kefalides, N. A. and Myers, J. C. (1985)
"Restricted homology between human $\alpha 1$ type IV and other procollagen chains"
Proc. Natl. Acad. Sci. USA 82, 3649-3653
- Burden, S. J., Sargent, P. B. and McMahan, U. J. (1979)
"Acetylcholine receptors in regenerating muscle accumulate at original synaptic sites in the absence of the nerve"
J. Cell Biol. 82, 412-425

- Capetanaki, Y. G., Ngai, J., Flytzanis, C. N. and Lazarides, E. (1983)
"Tissue-Specific Expression of Two mRNA Species Transcribed from a Single Vimentin Gen"
Cell 35, 411-420
- Carlin, B., Jaffe, R., Bender B. and Chung, A. E. (1981)
"Entactin, a Novel Basal-Lamina-Associated Sulfated Glycoprotein"
J. Biol. Chem. 256, 5209-5214
- Chen, E. Y. and Seeburg, P. H. (1985)
"Supercoil Sequencing: A Fast and Simple Method for Sequencing Plasmid DNA"
DNA 4, 165-170
- Cornish-Bowden, A. (1985)
"Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984
Nucl. Acid Res. 13, 3021-3030
- Courtoy, P. J., Timpl, R. and Farquhar, M. G. (1982)
"Comparative Distribution of Laminin, Type IV Collagen, and Fibronectin in the Rat Glomerulus"
J. Histochem. Cytochem. 30, 874-886
- Dean, D. C., Barr, J. F., Freytag, J. W. and Hudson, B. G. (1983)
"Isolation of Type IV Procollagen-like Polypeptides from Glomerular Basement Membrane"
J. Biol. Chem. 258, 590-596
- Devereux, J., Haeberli, P. and Smithies, O. (1984)
"A comprehensive set of sequence analysis programs for the VAX"
Nucl. Acids Res. 12, 387-395
- Dickson, L. A., Ninomiya, Y., Bernard, M. P., Pesciotta, D. M., Parsons, J., Green, G., Eikenberry, E. F., de Crombrughe, B., Vogeli, G., Pastan, I., Fietzek, P. P. and Olson, E. R. (1981)
"The exon/intron structure of the 3' region of the pro α 2(I) collagen gene"
J. Biol. Chem. 256, 8407-8415
- Dixit, S. N. and Kang, A. H. (1979)
"Anterior Lens Capsule Collagens: Cyanogen Bromide Peptides of the C Chain"
Biochemistry 18, 5686-5692
- Dixit, S. N. (1980)
"Type IV Collagens. Isolation and characterization of two structurally distinct collagen chains from bovine kidney cortices"
Eur. J. Biochem. 106, 563-570
- Dixit, S. N., Mainardi, C. L., Beachey, E. H. and Kang, A. H. (1983)
"7S Domain Constitutes the Amino-Terminal End of Type IV Collagen: an Immunochemical Study"
Coll. Rel. Res. 3, 263-270

- Doege, K. J. and Fessler, J. H. (1986)
"Folding of Carboxyl Domain and Assembly of Procollagen I"
J. Biol. Chem. 261, 8924-8935
- Dziadek, M. and Timpl, R. (1985)
"Expression of Nidogen and Laminin in Basement Membranes during Mouse Embryogenesis and in Teratocarcinoma Cells"
Dev. Biol. 111, 372-382
- Edgar, D., Timpl, R. and Thoenen, H. (1984)
"The heparin-binding domain of laminin is responsible for its effects on neurite outgrowth and neuronal survival"
EMBO J. 3, 1463-1468
- Farquhar, M. G. (1982)
"The Glomerular Basement Membrane - A Selective Macromolecular Filter"
In Hay, E. D. (ed.): "The Cell Biology of the Extracellular Matrix", New York, Plenum Press, 335-378
- Fessler, I., Timpl, R. and Fessler, J. H. (1981)
"Assembly and processing of procollagen type III in chick embryo blood vessels"
J. Biol. Chem. 256, 2531-2537
- Fessler, L. I. and Fessler, J. H. (1982)
"Identification of the Carboxyl Peptides of Mouse Procollagen IV and its Implications for the Assembly and Structure of Basement Membrane Procollagen"
J. Biol. Chem. 257, 9804-9810
- Fitzgerald, M. and Shenk, T. (1981)
"The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs"
Cell 24, 251-260
- Focht, R. J. and Adams, S. L. (1984)
"Tissue Specificity of Type I Collagen Gene Expression Is Determined at Both Transcriptional and Post-Transcriptional Levels"
Mol. Cell. Biol. 4, 1843-1852
- Fujiwara, S., Wiedemann, H., Timpl, R., Lustig, A. and Engel, J. (1984)
"Structure and interactions of heparan sulfate proteoglycans from a mouse tumor basement membrane"
Eur. J. Biochem. 143, 145-157
- Fuller, F. and Boedtker, H. (1981)
"Sequence determination and analysis of the 3' region of chicken pro α 1 and pro α 2 collagen messenger ribonucleic acids including the carboxyterminal propeptide sequences"
Biochemistry 20, 996-1006
- Gay, S. and Miller, E. J. (1979)
"Characterization of lens capsule collagen: evidence for the presence of two unique chains in molecules derived from major basement membrane structure"
Archs. Biochem. Biophys. 198, 370-378

- Gelman, R. A., Blackwell, J., Kefalides, N. A. and Tomicsek, E. (1976)
"Thermal stability of basement membrane collagen"
Biochim. Biophys. Acta 427, 492-496
- George, D. G., Orcutt, B. C., Dayhoff, M. O. and Barker, W. C. (1982)
in "NBRF-Reports", National Biomedical Research Foundation, Georgetown
University, Medical Center, Washington
- Gilbert, W. (1978)
"Why genes in pieces?"
Nature 271, 501
- Glanville, R. W., Rauter, A. and Fietzek, P. P. (1979)
"Isolation and Characterization of a Native Placental Basement-Membrane Collagen
and its Component Alpha-Chains"
Eur. J. Biochem. 95, 383-389
- Glanville, R. W., Voss, T. and Kühn, K. (1982)
"A Comparison of the Flexibility of Molecules of Basement Membrane and
Interstitial Collagens"
in Kühn, K., Schoene, H. and Timpl, R. (eds.): "New Trends in Basement Membrane
Research", Raven Press, New York, 69-77
- Glanville, R. W., Qian, R., Siebold, B., Risteli, J. and Kühn, K. (1985)
"Amino acid sequence of the N-terminal aggregation and cross-linking region
(7S domain) of the $\alpha 1$ (IV) chain of human basement membrane collagen"
Eur. J. Biochem. 152, 213-219
- Gopinathan, K. P., Weymouth, L. A., Kunkel, T. A. and Loeb, L. A. (1979)
"Mutagenesis in vitro by DNA polymerase from an RNA tumour virus"
Nature 278, 857-859
- Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981)
"Codon catalog usage is a genome strategy modulated for gene expressivity"
Nucl. Acids Res. 9, r43-r74
- Grosjean, H. and Fiers, W. (1982)
"Preferential codon usage in prokaryotic genes: the optimal codon-anticodon
interaction energy and the selective codon usage in efficiently expressed genes"
Gene 18, 199-209
- Hanahan, D. (1983)
"Studies on Transformation of *Escherichia coli* with Plasmids"
J. Mol. Biol. 166, 557-580
- Hanahan, D. (1985)
"Techniques for transformation of *E. coli*"
in Glover, D. M. (ed.): "DNA Cloning I", IRL Press, Oxford, 109-135
- Hofmann, H., Fietzek, P. P. and Kühn, K. (1980)
"Comparative analysis of the sequences of the three collagen chains $\alpha 1$ (I), $\alpha 2$ (I)
and $\alpha 1$ (III). Functional and genetic aspects"
J. Mol. Biol. 141, 293-314

- Hofmann, H., Voss, T., Kühn, K. and Engel, J. (1984)
"Localization of Flexible Sites in Thread-Like Molecules from Electron Micrographs. Comparison of Interstitial, Basement Membrane and Intima Collagens"
J. Mol. Biol. 172, 325-343
- Huerre, C., Junien, C., Weil, D., Chu, M., Morabito, M., Van Cong, N., Myers, J. C., Foubert, C., Gross, M., Prockop, D., Boue, A., Kaplan, J., De La Chappelle, A. and Ramirez, F. (1982)
"Human type I pro collagen genes are located on different chromosomes"
Proc. Natl. Acad. Sci. USA 79, 6627-6630
- Hulmes, D. J. S., Miller, A., Parry, D. A. D., Piez, K. and Woodhead-Galloway, J. (1973)
"Analysis of the primary structure of collagen for the origins of molecular packing"
J. Mol. Biol. 79, 137-148
- Ikemura, T. (1981)
"Correlation between the Abundance of *Escherichia coli* Transfer RNAs and the Occurrence of the Respective Codons in its Protein Genes: A Proposal for a Synonymous Codon Choice that is Optimal for the *E. coli* Translational System"
J. Mol. Biol. 151, 389-409
- Kanwar, Y. S. and Farquhar M. G. (1979)
"Anionic Sites in the Glomerular Basement Membrane. In Vivo and in Vitro Localization to the Laminar Rarae by Cationic Probes"
J. Cell Biol. 81, 137-153
- Kefalides, N. A. (1973)
"Structure and Biosynthesis of Basement Membranes"
Int. Rev. Connect. Tissue Res. 6, 63-104
- Kieny, M. P., Lathe, R. and Lecocq, J. P. (1983)
"New versatile cloning and sequencing vectors based on bacteriophage M13"
Gene 26, 91-99
- Kilchherr, E., Hofmann, H., Steigemann, W. and Engel, J. (1985)
"Structural Model of the Collagen-like Region of C1q Comprising the Kink Region of the Fibre-like Packing of the Six Triple Helices"
J. Mol. Biol. 186, 403-415
- Kozak, M. (1984)
"Compilation and analysis of the sequences upstream from the translational start site in eukaryotic mRNAs"
Nucl. Acids Res. 12, 857-872
- Kramer, J. M., Cox, G. M. and Hirsh, D. (1982)
"Comparisons of the Complete Sequences of Two Collagen Genes from *Caenorhabditis elegans*"
Cell 30, 599-606

- Kühn, K., Wiedemann, H., Timpl, R., Risteli, J., Dieringer, H., Voss, T. and Glanville, R. W. (1981)
"Macromolecular structure of basement membrane collagens. Identification of 7S collagen as a cross-linking domain of type-IV collagen"
FEBS Lett. 125, 123-128
- Kühn, K. (1984)
"Structural and Functional Domains of Collagen: A Comparison of the Protein with its 'Gene'"
Collagen Rel. Res. 4, 309-322
- Kurkinen, M., Barlow, D. P., Helfman, D. M., Williams, J. G. and Hogan, B. L. M. (1983)
"Isolation of cDNA clones for basal lamina components: type IV procollagen"
Nucl. Acids Res. 11, 6166-6208
- Kurkinen, M., Bernard, M. P., Barlow, D. P. and Chow, L. T. (1985)
"Characterization of 64-, 123- and 182-base-pair exons in the mouse $\alpha 2$ (IV) collagen gene"
Nature 317, 177-179
- Land, H., Grez, M., Hauser, H., Lindenmaier, W. and Schütz, G. (1981)
"5' terminal sequences of eucaryotic mRNA can be cloned with high efficiency"
Nucl. Acids Res. 9, 2251-2267
- Laurent, M., Martin, G. R. und Sobel, M. E.
"Cell free translation products of basement membrane RNA from the EHS tumor"
Biochim. Biophys. Acta, eingereicht
- Leivo, I., Alitalo, K., Risteli, L., Vaheri, A., Timpl, R. and Wartiovaara, J. (1982)
"Basal lamina glycoproteins laminin and type IV collagen are assembled into a fine-fibred matrix in cultures of a teratocarcinoma-derived endodermal cell line"
Exp. Cell Res. 137, 15-23
- Lehmann, J. M., Speers, W. C., Swartzendruber, D. E. and Pierce, G. B. (1974)
"Neoplastic differentiation: characterization of cell lines derived from a murine teratocarcinoma"
J. Cell Physiol. 84, 13-28
- Lozano, G., Ninomya, Y., Thompson, H. and Olsen, B. R. (1985)
"A distinct class of vertebrate collagen genes encodes chicken type IX collagen polypeptides"
Proc. Natl. Acad. Sci. USA 82, 4050-4054
- Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982)
"Molecular Cloning. A Laboratory Manual"
Cold Spring Harbor Laboratory, New York
- Mather, E. L., Nelson, K. J., Haimovich, J. and Perry, R. P. (1984)
"Mode of regulation of immunoglobulin μ - and δ -chain expression varies during B-lymphocyte maturation"
Cell 36, 329-338

- Maxam, A. M. and Gilbert, W. (1980)
"Sequencing End-Labeled DNA with Base-Specific Chemical Cleavages"
Meth. Enzymol. 65, 499-560
- McDevitt, M. A., Imperiale, M. J., Ali, H. and Nevins, J. R. (1984)
"Requirement of a Downstream Sequence for Generation of a Poly(A) Addition Site"
Cell 37, 993-999
- Meinkoth, J. and Wahl, G. (1984)
"Hybridization of Nucleic Acids Immobilized on Solid Supports"
Anal. Biochem. 138, 267-284
- Mercer, J. F. B. and Wake S. A. (1985)
"An analysis of the rate of metallothionein mRNA poly(A)-shortening using RNA blot hybridization"
Nucl. Acids Res. 13, 7929-7943
- Merlino, G. T., McKeon, C., de Crombrughe, B. and Pastan, I. (1983)
"Regulation of the Expression of Genes Encoding Types I, II, and III Collagen during Chick Embryonic Development"
J. Biol. Chem. 258, 10041-10048
- Merril, C. R., Goldman, D., Sedman, S. A. and Ebert, M. H. (1981)
"Ultrasensitive Stain for Proteins in Polyacrylamide Gels Shows Regional Variation in Cerebrospinal Fluid Proteins"
Science 211, 1437-1438
- Monson, J. M. and McCarthy, B. J. (1981)
"Identification of a Balb/c Mouse Pro α 1(I) Procollagen Gene: Evidence for Insertions or Deletions in Gene Coding Sequences"
DNA 1, 59-69
- Monson, J. M., Natzle, J., Friedman, J. and McCarthy, B. J. (1982)
"Expression and novel structure of a collagen gene in *Drosophila*"
Proc. Natl. Acad. Sci. USA 79, 1761-1765
- Myers, J. C., Brinker, J. M., Kefalides, N. A., Rosenblom, J., Wang, S. and Gudas, L. J. (1986)
"Discrimination among multiple AATAAA sequences correlates with interspecies conservation of select 3' untranslated nucleotides"
Nucl. Acids Res. 14, 4499-45017
- Nevins, J. R. (1983)
"The pathway of eukaryotic mRNA formation"
Ann. Rev. Biochem. 52, 441-466
- Oberbäumer, I., Wiedemann, H., Timpl, R. and Kühn, K. (1982)
"Shape and assembly of type IV procollagen obtained from cell culture"
EMBO J. 1, 805-810

- Oberbäumer, I., Laurent, M., Schwarz, U., Sakurai, Y., Yamada, Y., Vogeli, G., Voss, T., Siebold, B., Glanville, W. und Kühn, K. (1985)
"Amino acid sequence of the non-collagenous globular domain (NC1) of the $\alpha 1$ (IV) chain of basement membrane collagen as derived from complementary DNA"
Eur. J. Biochem. 147, 217-224
- Oberbäumer, I. (1986)
"New pUC-derived expression vectors for rapid construction of cDNA libraries"
Gene 49, 81-91
- Okayama, H. and Berg, P. (1982)
"High-Efficiency Cloning of Full-Length cDNA"
Mol. Cell. Biol. 2, 161-170
- Orkin, R. W., Gehron, P., McGoodwin, E. B., Martin, G. R., Valentine, T. and Swarm, R. (1977)
"Murine tumor producing a matrix of basement membrane"
J. Exp. Med. 145, 204-220
- Paulsson, M., Dziadek, M., Suchanek, C., Huttner, W. B. and Timpl, R. (1985)
"Nature of sulphated macromolecules in mouse Reichert's membrane"
Biochem. J. 231, 571-579
- Pfarr, D. S., Rieser, L. A., Woychik, R. P., Rottman, F. M., Rosenberg, M. and Reff, M. E. (1986)
"Differential Effects of Polyadenylation Regions on Gene Expression in Mammalian Cells"
DNA 5, 115-122
- Pihlajaniemi, T., Tryggvason, K., Myers, J. C., Kurkinen, M., Lebo, R., Cheung, M., Prockop, D. J. and Boyd, C. D. (1985)
"cDNA Clones Coding for the Pro- $\alpha 1$ (IV) Chain of Human Type IV Procollagen Reveal an Unusual Homology of Amino Acid Sequences in Two Halves of the Carboxyl-terminal Domain"
J. Biol. Chem. 260, 7681-7687
- Proudfoot, N. J. and Brownlee, G. G. (1976)
"3' non-coding region sequences in eukaryotic mRNA"
Nature 263, 211-214
- Raghow, R., Gossage, D. and Kang, A. H. (1986)
"Pretranslational Regulation of Type I Collagen, Fibronectin, and a 50-Kilodalton Noncollagenous Extracellular Protein by Dexamethasone in Rat Fibroblasts"
J. Biol. Chem. 261, 4677-4684
- Risteli, J., Bächinger, H. P., Furthmayr, H., Engel, J. and Timpl, R. (1980)
"7S Collagen: Characterization of an Unusual Basement Membrane Structure"
Eur. J. Biochem. 108, 239-250
- Rüther, U., Koenen, M., Otto, K. and Müller-Hill, B. (1981)
"pUR222, a vector for cloning and rapid chemical sequencing of DNA"
Nucl. Acids Res. 9, 4087-4099

- Sakakibara, S., Inonye, K., Shudo, K., Kishida, Y., Kobayashi, Y. and Prockop, D. J. (1973)
"Synthesis of (Pro-Hyp-Gly)_n of defined molecular weights. Evidence for the stabilization of collagen triple helix by hydroxyproline"
Biochim. Biophys. Acta 303, 198-202
- Sakashita, S., Engvall, E. and Ruoslahti, E. (1980)
"Basement membrane glycoprotein laminin binds to heparin"
FEBS Lett. 116, 243-246
- Sakurai, Y., Sullivan, M. and Yamada, Y. (1986)
" $\alpha 1$ Type IV Collagen Gene Evolved Differently from Fibrillar Collagen Genes"
J. Biol. Chem. 261, 6654-6657
- Sanes, J. R., Marshall, L. M. and McMahan, U. J. (1978)
"Reinnervation of muscle fiber basal lamina after removal of myofibers"
J. Cell Biol. 78, 176-198
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. and Roe, B. A. (1980)
"Cloning in Single-Stranded Bacteriophage as an Aid to Rapid DNA Sequencing"
J. Mol. Biol. 143, 161-178
- Schuppan, D., Glanville, R. W. and Timpl, R. (1982)
"Covalent Structure of Mouse Type-IV Collagen. Isolation, Order and Partial Amino-Acid Sequence of Cyanogen-Bromide and Tryptic Peptides of Pepsin Fragment P1 from the $\alpha 1$ (IV) Chain"
Eur. J. Biochem. 123, 505-512
- Schuppan, D., Glanville, R. W., Timpl, R., Dixit, S. N. and Kang, A. H. (1984)
"Sequence comparison of pepsin-resistant segments of basement-membrane collagen alpha-1 (IV) chains from bovine lens capsule and mouse-tumor"
Biochem. J. 220, 227-233
- Schwarz, U., Schuppan, D., Oberbäumer, I., Glanville, R. W., Deutzmann, R., Timpl, R. and Kühn, K. (1986)
"Structure of mouse type IV collagen. Amino-acid sequence of the C-terminal 511-residue-long triple-helical segment of the $\alpha 2$ (IV) chain and its comparison with the $\alpha 1$ (IV) chain"
Eur. J. Biochem. 157, 49-56
- Sharp, P. M., Tuohy, T. M. F. and Mosurski, K. R. (1986)
"Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes"
Nucl. Acids Res. 14, 5125-5143
- Siebold, B. (1986)
"Struktur der N- und C-terminalen Domänen des Basalmembranen Kollagens"
Dissertation, Universität München
- Soininen, R., Tikka, L., Chow, L., Fihlajaniemi, T., Kurkinen, M., Prockop, D. J., Boyd, C. D. and Tryggvason, K. (1986)
"Large introns in the 3' end of the gene for the pro $\alpha 1$ (IV) chain of human basement membrane collagen"
Proc. Natl. Acad. Sci. USA 83, 1568-1572

- Solomon, E., Hiorns, L., Dagleish, R., Tolstoshev, P., Crystal, R. and Sykes, B. (1983)
"Regional localization of the human $\alpha 2(I)$ collagen gene on chromosome 7, by molecular hybridization"
Cytogenet. Cell. Genet. 35, 64-66
- Stenman, S. and Vaheri, A. (1978)
"Distribution of a major connective tissue protein, fibronectin, in normal human tissues"
J. Exp. Med. 147, 1054-1064
- Stephenson, M. L., Krane, S. M., Amento, E. P., McCroskery, P. A. and Byrne, M. (1985)
"Immune interferon inhibits collagen synthesis by rheumatoid synovial cells associated with decreased levels of the procollagen mRNAs"
FEBS Letters 180, 43-49
- Strickland, S., Smith, K. K. and Marotti, K. R. (1980)
"The hormonal induction of differentiation in teratocarcinoma stem cells: generation of parietal endoderm by retinoic acid and dibutyryl cAMP"
Cell 21, 347-355
- Stryer, L. (1981)
"Biochemistry"
W. H. Freeman and Company, San Francisco
- Szostak, J. W., Stiles, J. I., Tye, B. K., Chin, P., Sherman, F. and Wu, R. (1979)
"Hybridization with synthetic oligonucleotides"
Meth. Enzymol. 68, 419-428
- Terranova, V. P., Rohrbach, D. H. and Martin, G. R. (1980)
"Role of laminin in the attachment of PAM 212 (epithelial) cells to basement membrane collagen"
Cell 22, 719-726
- Timpl, R., Bruckner, P. and Fietzek, P. P. (1979)
"Characterization of Pepsin Fragments of Basement Membrane Collagen Obtained from a Mouse Tumor"
Eur. J. Biochem. 95, 255-263
- Timpl, R., Rohde, H., Gehron Robey, P., Rennard, S. I., Foidart, J. M. and Martin, G. R. (1979)
"Laminin - a Glycoprotein from Basement Membranes"
J. Biol. Chem. 254, 9933-9937
- Timpl, R., Wiedemann, H., Van Delden, V., Furthmayr, H. and Kühn, K. (1981)
"A Network Model for the Organization of Type IV Collagen Molecules in Basement Membranes"
Eur. J. Biochem. 120, 203-211
- Timpl, R., Dziadek, M., Fujiwara, S., Nowack, H. and Wick, G. (1983)
"Nidogen, a New Self-Aggregating Basement Membrane Protein"
Eur. J. Biochem. 137, 455-465

- Timpl, R. and Dziadek, M. (1986)
"Structure, Development and Molecular Pathology of Basement Membranes"
Int. Rev. Exp. Path. 29, 1-112
- Trüb, B., Gröbli, B., Spiess, M., Odermatt, B. F. and Winterhalter, K. H. (1982)
"Basement Membrane (Type IV) Collagen Is a Heteropolymer"
J. Biol. Chem. 257, 5239-5245
- Tryggvason, K., Gehron Robey, P. and Martin, G. R. (1980)
"Biosynthesis of Type IV Procollagens"
Biochemistry 19, 1284-1289
- Tsao, S. G. S., Brunk, C. F. and Pearlman, R. E. (1983)
"Hybridization of Nucleic Acids Directly in Agarose Gels"
Anal. Biochem. 131, 365-372
- Vieria, J. and Messing, J. (1982)
"The pUC plasmids, a M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers"
Gene 19, 259-268
- Vogeli, G., Horn, E., Laurent, M. and Nath, P. (1985)
"Recombinant DNA Techniques: Storage and Screening of cDNA Libraries with Large Numbers of Individual Colonies from Initial Transformations"
Analyt. Biochem. 151, 442-444
- Vracko, R. (1982)
"The Role of Basal Lamina in Maintenance of Orderly Tissue Structure"
in Kühn, K., Schoene, H. and Timpl, R. (eds.): "New Trends in Basement Membrane Research", New York, Raven Press, 1-7
- Walker, F. (1972)
"Basement membrane turnover in the rat"
J. Pathol. 107, 119-121
- Wallace, R. B., Shaffer, J., Murphy, R.F., Bonner, J., Hirose, T. and Itakura, K. (1979)
"Hybridization of synthetic oligodeoxyribonucleotides to ϕ x 174 DNA: the effect of single base pair mismatch"
Nucl. Acids Res. 6, 3543-3557
- Wartiovaara, J., Leivo, I. and Vaheri, A. (1979)
"Expression of the cell surface-associated glycoprotein, fibronectin, in the early mouse embryo"
Dev. Biol. 69, 247-
- Weber, S., Engel, J., Wiedemann, H., Glanville, R. W. and Timpl, R. (1984)
"Subunit structure and assembly of the globular domain of basement membrane collagen"
Eur. J. Biochem. 139, 401-410
- Wilimzig, M. (1985)
"LiCl-boiling method for plasmid mini-preps"
Trends Genet. 1, 158

Wood, W. I., Gitschier, J., Lasky, L. A. and Lawn, R. M. (1985)

"Base composition-independent hybridization in tetramethylammonium chloride: A method for oligonucleotide screening of highly complex gene libraries"
Proc. Natl. Acad. Sci. USA 82, 1585-1588

Yaffe, D., Nudel, U., Mayer, Y. and Neumann, S. (1985)

"Highly conserved sequences in the 3' untranslated region of mRNAs coding for homologous proteins in distantly related species"
Nucl. Acids Res. 13, 3723-3737

Yamada, K. M. (1983)

"Cell surface interactions with extracellular materials"
Ann. Rev. Biochem. 52, 761-799

Yamada, Y., Kühn, K. and de Crombrugghe B. (1983)

"A conserved nucleotide sequence, coding for a segment of the C-propeptide, is found at the same location in different collagen genes"
Nucl. Acids Res. 11, 2733-2744

Yanisch-Perron, C., Viera, J. and Messing, J. (1985)

"Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors"
Gene 33, 103-119

Yacita, H., Foidart, J. M. and Katz, S. I. (1978)

"Localization of the collagenous component in skin basement membrane"
J. Invest. Dermatol. 70, 191-193

ABKÜRZUNGEN

bp	Basenpaar
BSA	Rinderserumalbumin
Ci	Curie
cpm	"counts" pro Minute
dNTP	Desoxynukleotidtriphosphat
ddNTP	2',3'-Didesoxynukleotidtriphosphat
DMS	Dimethylsulfat
DMSO	Dimethylsulfoxid
DTT	Dithiothreitol
EDTA	Ethylendiaminotetraessigsäure
h	Stunde
IPTG	Isopropyl- β -D-Thiogalactopyranosid
kb	Kilobasen
kd	Kilodalton
min	Minuten
OD	optische Dichte
PEG	Polyethylenglykol
SDS	Natriumdodecylsulfat
sec	Sekunde
TeMED	N,N,N',N'-Tetramethylethyldiamin
Tris	Trishydroxymethylaminomethan
u	Enzymeinheit (unit)
Upm	Umdrehungen pro Minute
X-Gal	5-Brom-4-Chlor-3-Indolyl- β -D-Galaktopyranosid

Danksagung

Am Ende dieser Arbeit gilt mein besonderer Dank Herrn Prof. Dr. K. Kühn für die Aufgabenstellung und Betreuung der Arbeit.

Zu großem Dank bin ich Frau Dr. I. Oberbäumer verpflichtet, die mich mit den molekularbiologischen Methoden und Techniken vertraut gemacht hat und mir jederzeit mit Rat und Hilfe zur Seite stand.

Bei meinem "Mitstreiter" D. Brazel und allen anderen Mitarbeitern möchte ich mich für die Hilfs- und Diskussionsbereitschaft, sowie das angenehme Arbeitsklima bedanken.

Herrn Prof. Dr. W. Bandlow danke ich für die Betreuung der Arbeit und das entgegengebrachte Interesse.

Für die in jeder Hinsicht gewährte Unterstützung danke ich meinem Mann, meinen Eltern und meinen Großeltern.

Lebenslauf

Ursula Magdolen, geb. Schwarz

geboren am 4. 2. 1957 in München

1963 - 1967	Volksschule in Putzbrunn bei München
1967 - 1975	Besuch des Luisengymnasiums in München, Neusprachlicher Zweig
1975	Abitur
1976 - 1981	Studium der Biologie an der Ludwig-Maximilians- Universität in München
1981 - 1983	Diplomarbeit am Genetischen Institut der Universität in München bei Prof. Dr. W. Bandlow Studienabschluß mit Diplom in Biologie
1983	Beginn der Dissertationsarbeit am Max-Planck-Institut für Biochemie bei Prof. Dr. K. Kühn
1986	Eheschließung mit Viktor Magdolen

